VISIT…

LANZAROTE
Caliente.com

A CRYPTOGRAPHIC SERIES

30

MILITARY CRYPTANALYSIS
PART I

With New Added Problems For The Student

	Pages
Introductory remarks	1- 6
Fundamental principles	7- 10
Frequency distributions	11- 17
Fundamental uses of the uniliteral frequency distribution	18- 26
Uniliteral substitution with standard cipher alphabets	27- 39
Uniliteral substitution with mixed cipher alphabets	40- 58
Multiliteral substitution with single-equivalent cipher alphabets	59- 62
Multiliteral substitution with multiple-equivalent cipher alphabets	63- 69
Polygraphic substitution systems	70- 98
Concluding remarks	99-104
Appendix	105-135
Index	137-142
Analytical key for cryptanalysis	142
PROBLEMS	143-149

by

William F. Friedman

From Aegean Park Press

With New Added Problems For The Student 1
FOREWORD 3
MILITARY CRYPTANALYSIS, PART I. 4
SECTION I 5
  INTRODUCTORY REMARKS 5
SECTION II 11
  FUNDAMENTAL PRINCIPLES 11
SECTION III 15
  FREQUENCY DISTBIBUTIONS 15
SECTION IV 22
  FUNDAMENTAL USES OF THE UNILITERAL 22
SECTION V 31
  UNILITERAL SUBSTITUTION WITH STANDARD 31
SECTION VI 44
  UNILITERAL SUBSTITUTION WITH MIXED CIPHER 44
SECTION VII 63
  MULTILITERAL SUBSTITUTION WITH SINGLE-EQUIV 63
SECTION VIII 67
  MULTHJTEUL SUBSTITUTION WITH MULTIPLE-EQUIV... 67
SECTION IX 74
  POLYGRAPHIC SUBSTITUTION SYSTEMS 74
SECTION X 103
  CONCLUDING REMARKS 103
APPENDIX I 109
INDEX 140
PROBLEMS 146
BOOKS IN TEH CRYPTOGRAPHIC SERIES 153

© 1980 Aegean Park Press

ISBN: 0-89412-044-1

Published by AEGEAN PARK PRESS
P.O. Box 2837, Laguna Hills, California 92653
(714)586-8811

Manufactured in the United States of America

FOREWORD

We are particularly pleased to add this book, recently declassified by the U.S. Government, MILITARY CRYPTANALYSIS, Part I, to our Cryptographic Series. Written by William F. Friedman, considered by many to be the father of modern cryptology, this classic treatise thoroughly and in a scientific manner discusses the cryptanalysis of Monoalphabetic Substitution Cipher Systems.

To make this text even more valuable to the student, we have added at the end of the text, beginning on page 143, a series of problems and questions, all carefully constructed, which in large part are keyed to the order in which the material is presented in the text. Many hours have been spent in devising these problems, all of which of course are completely new. The student is urged, incidentally, to solve these problems, for in the final analysis, one learns best by doing!

As with all books in the Cryptographic Series, we would like to have any comments you might have concerning this book. If you are able to successfully solve all the problems, we would like you to tell us that too.

September 1980 AEGEAN PARK PRESS

MILITARY CRYPTANALYSIS, PART I. MONOALPHABETIC SUBSTITUTION SYSTEMS

Section		Paragraphs	Pages
I.	Introductory remarks	1– 3	1– 6
II.	Fundamental principles	4– 8	7–10
III.	Frequency distributions	9–11	11–17
IV.	Fundamental uses of the uniliteral frequency distribution	12–16	18–26
V.	Uniliteral substitution with standard cipher alphabets	17–22	27–39
VI.	Uniliteral substitution with mixed cipher alphabets	23–34	40–58
VII.	Multiliteral substitution with single-equivalent cipher alphabets	35–36	59–62
VIII.	Multiliteral substitution with multiple-equivalent cipher alphabets	37–40	63–69
IX.	Polygraphic substitution systems	41–46	70–98
X.	Concluding remarks	47–50	99–104
XI.	Appendix		105–120

SECTION I

INTRODUCTORY REMARKS

Paragraph
Scope of this text 1
Mental equipment necessary for cryptanalytic work 2
Validity of results of cryptanalysis 3

1. **Scope of this text.**—*a.* It is assumed that the student has studied the two preceding texts written by the same author and forming part of this series, viz, *Elementary Military Cryptography*, and *Advanced Military Cryptography*. These texts deal exclusively with *cryptography* as defined therein; that is, with the various types of ciphers and codes, their principles of construction, and their employment in cryptographing and decryptographing messages. Particular emphasis was placed upon such means and methods as are practicable for military usage. It is also assumed that the student has firmly in mind the technically precise, special nomenclature employed in those texts, for the terms and definitions therein will all be used in the present text, with essentially the same significances. If this is not the case, it is recommended that the student review his preceding work, in order to regain a familiarity with the specific meanings assigned to the terms used therein. There will be no opportunity herein to repeat this information and unless he understands clearly the significance of the terms employed, his progress will be retarded.

b. This text constitutes the first of a series of texts on *cryptanalysis*. Although most of the information contained herein is applicable to cryptograms of various types and sources, special emphasis will be laid upon the principles and methods of solving military cryptograms. Except for an introductory discussion of fundamental principles underlying the science of cryptanalytics, this first text in the series will deal solely with the principles and methods for the analysis of monoalphabetic substitution ciphers. Even with this limitation it will be possible to discuss only a few of the many variations of this one type; but with a firm grasp upon the general principles no difficulties should be experienced with any variations that may be encountered.

c. This and some of the succeeding texts will deal only with elementary types of cipher systems not because they may be encountered in military operations but because their study is essential to an understanding of the principles underlying the solution of the modern, very much more complex types of ciphers and codes that are likely to be employed by the larger governments today in the conduct of their military affairs in time of war.

d. All of this series of texts will deal only with the solution of visible secret writing. At some future date, texts dealing with the solution of invisible secret writing, and with secret signalling systems, may be prepared.

2. **Mental equipment necessary for cryptanalytic work.**—*a.* Captain Parker Hitt, in the first United States Army manual[1] dealing with cryptography, opens the first chapter of his valuable treatise with the following sentence:

> Success in dealing with unknown ciphers is measured by these four things in the order named: perseverance, careful methods of analysis, intuition, luck.

---

[1] Hitt, Capt. Parker, *Manual for the Solution of Military Ciphers.*

These words are as true today as they were then. There is no royal road to success in the solution of cryptograms. Hitt goes on to say:

> Cipher work will have little permanent attraction for one who expects results at once, without labor, for there is a vast amount of purely routine labor in the preparation of frequency tables, the rearrangement of ciphers for examination, and the trial and fitting of letter to letter before the message begins to appear.

The present author deems it advisable to add that the kind of work involved in solving cryptograms is not at all similar to that involved in solving "cross-word puzzles", for example. The wide vogue the latter have had and continue to have is due to the appeal they make to the quite common instinct for mysteries of one sort or another; but in solving a cross-word puzzle there is usually no necessity for performing any preliminary labor, and palpable results become evident after the first minute or two of attention. This successful start spurs the cross-word "addict" on to complete the solution, which rarely requires more than an hour's time. Furthermore, cross-word puzzles are all alike in basic principle and once understood, there is no more to learn. Skill comes largely from the embellishment of one's vocabulary, though, to be sure, constant practice and exercise of the imagination contribute to the ease and rapidity with which solutions are generally reached: In solving cryptograms, however, many principles must be learned, for there are many different systems of varying degrees of complexity. Even some of the simpler varieties require the preparation of tabulations of one sort or another, which many people find irksome; moreover, it is only toward the very close of the solution that results in the form of intelligible text become evident. Often, indeed, the student will not even know whether he is on the right track until he has performed a large amount of preliminary "spade work" involving many hours of labor. Thus, without at least a willingness to pursue a fair amount of theoretical study, and a *more than average amount of patience and perseverance*, little skill and experience can be gained in the rather difficult art of cryptanalysis. General Givierge, the author of an excellent treatise on cryptanalysis, remarks in this connection: [2]

> The cryptanalyst's attitude must be that of William the Silent: No need to hope in order to undertake, nor to succeed in order to persevere.

b. As regards Hitt's reference to careful methods of analysis, before one can be said to be a cryptanalyst worthy of the name it is necessary that one should have firstly a sound knowledge of the basic principles of cryptanalysis, and secondly, a long, varied, and active *practical* experience in the successful application of those principles. It is not sufficient to have read treatises on this subject. One month's actual practice in solution is worth a whole year's mere reading of theoretical principles. An exceedingly important element of success in solving the more intricate ciphers is the possession of the rather unusual mental faculty designated in general terms as the power of inductive and deductive reasoning. Probably this is an inherited rather than an acquired faculty; the best sort of training for its emergence, if latent in the individual, and for its development is the study of the natural sciences, such as chemistry, physics, biology, geology, and the like. Other sciences such as linguistics and philology are also excellent. Aptitude in mathematics is quite important, more especially in the solution of ciphers than of codes.

c. An active imagination, or perhaps what Hitt and other writers call *intuition*, is essential, but mere imagination uncontrolled by a judicious spirit will more often be a hindrance than a help. In practical cryptanalysis the imaginative or intuitive faculties must, in other words, be guided by good judgment, by practical experience, and by as thorough a knowledge of the general situation or extraneous circumstances that led to the sending of the cryptogram as is possible to obtain. In this respect the many cryptograms exchanged between correspondents whose identities and general affairs, commercial, social, or political, are known are far more readily solved than are isolated cryptograms exchanged between unknown correspondents, dealing with unknown subjects. It is obvious that in the former case there are good data upon which the intuitive powers of the cryptanalyst can be brought to bear, whereas in the latter case no such data are available. Consequently, in the absence of such data, no matter how good the imagination and intuition of the cryptanalyst, these powers are of no particular service to him. Some writers, however, regard the intuitive spirit as valuable from still another viewpoint, as may be noted in the following:[3]

---

[2] Givierge, Général Marcel, *Cours de Cryptographie*, Paris, 1925, p. 301.

> Intuition, like a flash of lightning, lasts only for a second. It generally comes when one is tormented by a difficult decipherment and when one reviews in his mind the fruitless experiments already tried. Suddenly the light breaks through and one finds after a few minutes what previous days of labor were unable to reveal.

This, too, is true, but unfortunately there is no way in which the intuition may be summoned at will, when it is most needed.[4] There are certain authors who regard as indispensable the possession of a somewhat rare, rather mysterious faculty that they designate by the word "flair", or by the expression "cipher brains." Even so excellent an authority as General Givierge,[5] in referring to this mental facility, uses the following words: "Over and above perseverance and this aptitude of mind which some authors consider a special gift, and which they call intuition, or even, in its highest manifestation, clairvoyance, cryptographic studies will continue more and more to demand the qualities of orderliness and memory." Although the present author believes a special aptitude for the work is essential to cryptanalytic success, he is sure there is nothing mysterious about the matter at all. Special aptitude is prerequisite to success in all fields of endeavor. There are, for example, thousands of physicists, hundreds of excellent ones, but only a handful of world-wide fame. Should it be said, then, that a physicist who has achieved very notable success in his field has done so because he is the fortunate posessor of a *mysterious* faculty? That he is fortunate in possessing a special aptitude for his subject is granted, but that there is anything mysterious about it, partaking of the nature of clairvoyance (if, indeed, the latter is a *reality*) is not granted. While the ultimate nature of any mental process seems to be as complete a mystery today as it has ever been, the present author would like to see the superficial veil of mystery removed from a subject that has been shrouded in mystery from even before the Middle Ages down to our own times. (The principal and easily understandable reason for this is that governments have always closely guarded cryptographic secrets and anything so guarded soon becomes "mysterious.") He would, rather, have the student approach the subject as he might approach any other science that can stand on its own merits with other sciences, because cryptanalytics, like other sciences, has a practical importance in human affairs. It presents to the inquiring mind an interest in its own right as a branch of knowledge; it, too, holds forth many difficulties and disappointments, and these are all the more keenly felt when the nature of these difficulties is not understood by those unfamiliar with the special circumstances that very often are the real factors that led to success in other cases. Finally, just as in the other sciences wherein many men labor long and earnestly for the true satisfaction and pleasure that comes from work well-done, so the mental pleasure that the successful cryptanalyst derives from his accomplishments is very often the only reward for much of the drudgery that he must do in his daily work. General Givierge's words in this connection are well worth quoting:[6]

---

[3] Lange et Soudart, *Traité de Cryptographie*, Librairie Félix Alcan, Paris, 1925, p. 104.

[4] The following extracts are of interest in this connection:

"The fact that the scientific investigator works 50 per cent of his time by non-rational means is, it seems, quite insufficiently recognized. There is without the least doubt an instinct for research, and often the most successful investigators of nature are quite unable to give an account of their reasons for doing such and such an experiment, or for placing side by side two apparently unrelated facts. Again, one of the most salient traits in the character of the successful scientific worker is the capacity for knowing that a point is proved when it would not appear to be proved to an outside intelligence functioning in a purely rational manner; thus the investigator feels that some proposition is true, and proceeds at once to the next set of experiments without waiting and wasting time in the elaboration of the formal proof of the point which heavier minds would need. Questionless such a scientific intuition may and does sometimes lead investigators astray, but it is quite certain that if they did not widely make use of it, they would not get a quarter as far as they do. Experiments confirm each other, and a false step is usually soon discovered. And not only by this partial replacement of reason by intuition does the work of science go on, but also to the born scientific worker—and emphatically they cannot be made—the structure of the method of research is as it were given, he cannot explain it to you, though he may be brought to agree *a posteriori* to a formal logical presentation of the way the method works".—Excerpt from Needham, Joseph, *The Sceptical Biologist*, London, 1929, p. 79.

"The essence of scientific method, quite simply, is to try to see how data arrange themselves into causal configurations. Scientific problems are solved by collecting data and by "thinking about them all the time." We need to look at strange things until, by the appearance of known configurations, they seem familiar, and to look at familiar things until we see novel configurations which make them appear strange. We must look at events until they become luminous. That is scientific method . . . Insight is the touchstone . . . The application of insight as the touchstone of method enables us to evaluate properly the role of imagination in scientific method. The scientific process is akin to the artistic process: it is a process of selecting out those elements of experience which fit together and recombining them in the mind. Much of this kind of research is simply a ceaseless mulling over, and even the physical scientist has considerable need of an armchair . . . Our view of scientific method as a struggle to obtain insight forces the admission that science is half art . . . Insight is the unknown quantity which has eluded students of scientific method".—Excerpts from an article entitled *Insight and Scientific Method*, by Willard Waller, in *The American Journal of Sociology*, Vol. XL, 1934.

[5] *Op. cit.*, p. 302.

> Some studies will last for years before bearing fruit. In the case of others, cryptanalysts undertaking them never get any result. But, for a cryptanalyst who likes the work, the joy of discoveries effaces the memory of his hours of doubt and impatience.

d. With his usual deft touch, Hitt says of the element of luck, as regards the role it plays in analysis:

> As to luck, there is the old miners' proverb: "Gold is where you find it."

The cryptanalyst is lucky when one of the correspondents whose ciphers he is studying makes a blunder that gives the necessary clue; or when he finds two cryptograms identical in text but in different keys in the same system; or when he finds two cryptograms identical in text but in different systems, and so on. The element of luck is there, to be sure, *but the cryptanalyst must be on the alert* if he is to profit by these lucky "breaks."

e. If the present author were asked to state, in view of the progress in the field since 1916, what elements might be added to the four ingredients Hitt thought essential to cryptanalytic success, he would be inclined to mention the following:

(1) A broad, general education, embodying interests covering as many fields of practical knowledge as possible. This is useful because the cryptanalyst is often called upon to solve messages dealing with the most varied of human activities, and the more he knows about these activities, the easier his task.

(2) Access to a large library of current literature, and wide and direct contacts with sources of collateral information. These often afford clues as to the contents of specific messages. For example, to be able instantly to have at his disposal a newspaper report or a personal report of events described or referred to in a message under investigation goes a long way toward simplifying or facilitating solution. Government cryptanalysts are sometimes fortunately situated in this respect, especially where various agencies work in harmony.

(3) Proper coordination of effort. This includes the organization of cryptanalytic personnel into harmonious, efficient teams of cooperating individuals.

---

[6] *Op. cit.*, p. 301.

(4) Under mental equipment he would also include the faculty of being able to concentrate on a problem for rather long periods of time, without distraction, nervous irritability, and impatience. The strain under which cryptanalytic studies are necessarily conducted is quite severe and too long-continued application has the effect of draining nervous energy to an unwholesome degree, so that a word or two of caution may not here be out of place. One should continue at work only so long as a peaceful, calm spirit prevails, whether the work is fruitful or not. But just as soon as the mind becomes wearied with the exertion, or just as soon as a feeling of hopelessness or mental fatigue intervenes, it is better to stop completely and turn to other activities, rest, or play. It is essential to remark that systematization and orderliness of work are aids in reducing nervous tension and irritability. On this account it is better to take the time to prepare the data carefully, rewrite the text if necessary, and so on, rather than work with slipshod, incomplete, or improperly arranged material.

(5) A retentive memory is an important asset to cryptanalytic skill, especially in the solution of codes. The ability to remember individual groups, their approximate locations in other messages, the associations they form with other groups, their peculiarities and similarities, saves much wear and tear of the mental machinery, as well as much time in looking up these groups in indexes.

f. It may be advisable to add a word or two at this point to prepare the student to expect slight mental jars and tensions which will almost inevitably come to him in the conscientious study of this and the subsequent texts. The present author is well aware of the complaint of students that authors of texts on cryptanalysis base much of their explanation upon their foreknowledge of the "answer"—which the student does not know while he is attempting to follow the solution with an unbiased mind. They complain too that these authors use such expressions as "obviously", "naturally", "of course", "it is evident that", and so on, when the circumstances seem not at all to warrant their use. There is no question but that this sort of treatment is apt to discourage the student, especially when the point elucidated becomes clear to *him* only after many hours' labor, whereas, according to the book, the author noted the weak spot at the first moment's inspection. The present author can only promise to try to avoid making the steps appear to be much more simple than they really are, and to suppress glaring instances of unjustifiable "jumping at conclusions." At the same time he must indicate that for pedagogical reasons in many cases a message has been consciously "manipulated" so as to allow certain principles to become more obvious in the illustrative examples than they ever are in practical work. During the course of some of the explanations attention will even be directed to cases of unjustified inferences. Furthermore, of the student who is quick in observation and deduction, the author will only ask that he bear in mind that if the elucidation of certain principles seems prolix and occupies more space than necessary, this is occasioned by the author's desire to carry the explanation forward in very short, easily-comprehended, and plainly-described steps, for the benefit of students who are perhaps a bit slower to grasp but who, once they understand, are able to retain and apply principles slowly learned just as well, if not better than the students who learn more quickly.

3. **Validity of results of cryptanalysis.**—Valid, or authentic cryptanalytic solutions cannot and do not represent "opinions" of the cryptanalyst. They are valid only so far as they are wholly objective, and are susceptible of demonstration and proof, employing authentic, objective methods. It should hardly be necessary (but an attitude frequently encountered among laymen makes it advisable) to indicate that the validity of the results achieved by any serious cryptanalytic studies on authentic material rests upon the same sure foundations and are reached by the same general steps as the results achieved by any other scientific studies; viz, observation, hypothesis, deduction and induction, and confirmatory experiment. Implied in the latter is the

possibility that two or more qualified investigators, each working independently upon the same material, will achieve identical (or practically identical) results. Occasionally a pseudo-cryptanalyst offers "solutions" which cannot withstand such tests; a second, unbiased, investigator working independently either cannot *consistently* apply the methods alleged to have been applied by the pseudo-cryptanalyst, or else, if he can apply them at all, the results (plain-text translations) are far different in the two cases. The reason for this is that in such cases it is generally found that the "methods" are not clear-cut, straightforward or mathematical in character. Instead, they often involve the making of judgments on matters too tenuous to measure, weigh, or otherwise subject to careful scrutiny. In such cases, the conclusion to which the unprejudiced observer is forced to come is that the alleged "solution" obtained by the first investigator, the pseudo-cryptanalyst, is purely subjective. In nearly all cases where this has happened (and they occur from time to time) there has been uncovered nothing which can in any way be used to impugn the integrity of the pseudo-cryptanalyst. The worst that can be said of him is that he has become a victim of a special or peculiar form of self-delusion, and that his desire to solve the problem, usually in accord with some previously-formed opinion, or notion, has over-balanced, or undermined, his judgment and good sense.[7]

---

[7] Specific reference can be made to the following typical "case histories":

Donnelly, Ignatius, *The Great Cryptogram.* Chicago, 1888.
Owen, Orville W., *Sir Francis Bacon's Cipher Story.* Detroit, 1895.
Gallup, Elizabeth Wells, *Francis Bacon's Biliteral Cipher.* Detroit, 1900.
Margoliouth, D. S., *The Homer of Aristotle.* Oxford, 1923.
Newbold, William Romaine, *The Cipher of Roger Bacon.* Philadelphia, 1928. (For a scholarly and complete demolition of Professor Newbold's work, see an article entitled *Roger Bacon and the Voynich MS*, by John M. Manly, in *Speculum*, Vol. *VI*, No. 3, July 1931.)
Arensberg, Walter Conrad, *The Cryptography of Shakespeare.* Los Angeles, 1922.
  The Shakespearean Mystery. Pittsburgh, 1928.
  The Baconian Keys. Pittsburgh, 1928.
Feely, Joseph Martin, *The Shakespearean Cypher.* Rochester, N. Y., 1931.
  Deciphering Shakespeare. Rochester, N. Y., 1934.

SECTION II

FUNDAMENTAL PRINCIPLES

Paragraph

The four basic operations in cryptanalysis 4

The determination of the language employed 5

The determination of the general system 6

The reconstruction of the specific key 7

The reconstruction of the plain text 8

4. The four basic operations in cryptanalysis.—*a.* The solution of practically every cryptogram involves four fundamental operations or steps:

(1) The determination of the language employed in the plain-text version.

(2) The determination of the general system of cryptography employed.

(3) The reconstruction of the specific key in the case of a cipher system, or the reconstruction, partial or complete, of the code book, in the case of a code system; or both, in the case of an enciphered code system.

(4) The reconstruction or establishment of the plain text.

b. These operations will be taken up in the order in which they are given above and in which they usually are performed in the solution of cryptograms, although occasionally the second step may precede the first.

5. The determination of the language employed.—*a.* There is not much that need be said with respect to this operation except that the determination of the language employed seldom comes into question in the case of studies made of the cryptograms of an organized enemy. By this is meant that during wartime the enemy is of course known, and it follows, therefore, that the language he employs in his messages will almost certainly be his native or mother tongue. Only occasionally nowadays is this rule broken. Formerly it often happened, or it might have indeed been the general rule, that the language used in diplomatic correspondence was not the mother tongue, but French. In isolated instances during the World War, the Germans used English when their own language could for one reason or another not be employed. For example, for a year or two before the entry of the United States into that war, during the time America was neutral and the German Government maintained its embassy in Washington, some of the messages exchanged between the Foreign Office in Berlin and the Embassy in Washington were cryptographed in English, and a copy of the code used was deposited with the Department of State and our censor. Another instance is found in the case of certain Hindu conspirators who were associated with and partially financed by the German Government in 1915 and 1916; they employed English as the language of their cryptographic messages. Occasionally the cryptograms of enemy agents may be in a language different from that of the enemy. But in general these are, as has been said, isolated instances; as a rule, the language used in cryptograms exchanged between members of large organizations is the mother tongue of the correspondents. Where this is not the case, that is, when cryptograms of unknown origin must be studied, the cryptanalyst looks for any indications on the cryptograms themselves which may lead to a conclusion as to the language employed. Address, signature, and plain-language words in the preamble or in the body of the text all come under careful scrutiny, as well as all extraneous circumstances connected with the manner in which the cryptograms were obtained, the person on whom they were found, or the locale of their origin and destination.

b. In special cases, or under special circumstances a clue to the language employed is found in the nature and composition of the cryptographic text itself. For example, if the letters K and W are entirely absent or appear very rarely in messages, it may indicate that the language is Spanish, for these letters are absent in the alphabet of that language and are used only to spell foreign words or names. The presence of accented letters or letters marked with special signs of one sort or another, peculiar to certain languages, will sometimes indicate the language used. The Japanese Morse telegraph alphabet and the Russian Morse telegraph alphabet contain combinations of dots and dashes which are peculiar to those alphabets and thus the interception of messages containing these special Morse combinations at once indicates the language involved. Finally, there are certain peculiarities of alphabetic languages which, in certain types of cryptograms, *viz,* pure transposition, give clues as to the language used. For example, the frequent digraph C H, in German, leads to the presence, in cryptograms of the type mentioned, of many isolated C's and H's; if this is noted, the cryptogram may be assumed to be in German.

c. In some cases it is perfectly possible to perform certain steps in cryptanalysis *before* the language of the cryptogram has been definitely determined. Frequency studies, for example, may be made and analytic processes performed without this knowledge, and by a cryptanalyst wholly unfamiliar with the language even if it has been identified, or who knows only enough about the language to enable him to recognize valid combinations of letters, syllables, or a few common words in that language. He may, after this, call to his assistance a translator who may not be a cryptanalyst but who can materially aid in making necessary assumptions based upon his special knowledge of the characteristics of the language in question. Thus, cooperation between cryptanalyst and translator results in solution.[1]

6. The determination of the general system.—*a.* Except in the case of the more simple types of cryptograms, the determination of the general system according to which a given cryptogram has been produced is usually a difficult, if not the most difficult, step in its solution. The reason for this is not hard to find.

b. As will become apparent to the student as he proceeds with his study, *in the final analysis, the solution of every cryptogram involving a form of substitution depends upon its reduction to monoalphabetic terms, if it is not originally in those terms.* This is true not only of ordinary substitution ciphers, but also of combined substitution-transposition ciphers, and of enciphered code. If the cryptogram must be reduced to monoalphabetic terms, the manner of its accomplishment is usually indicated by the cryptogram itself, by external or internal phenomena which become apparent to the cryptanalyst as he studies the cryptogram. If this is impossible, or too difficult the cryptanalyst must, by one means or another, discover how to accomplish this reduction, by bringing to bear all the special or collateral information he can get from all the sources at his command. If both these possibilities fail him, there is little left but the long, tedious, and often fruitless process of elimination. In the case of transposition ciphers of the more complex type, the discovery of the basic method is often simply a matter of long and tedious elimination of possibilities. For cryptanalysis has unfortunately not yet attained, and may indeed never attain, the precision found today in qualitative analysis in chemistry, for example, where the analytic process is absolutely clear cut and exact in its dichotomy. A few words in explanation of what is meant may not be amiss. When a chemist seeks to determine the identity of an unknown substance, he applies certain specific reagents to the substance and in a specific sequence. The first reagent tells him definitely into which of two primary classes the unknown substance falls. He then applies a second test with another specific reagent, which tells him again quite definitely into which of two secondary classes the unknown substance falls, and so on, until finally he has reduced the unknown substance to its simplest terms and has found out what it is. In striking contrast to this situation, cryptanalysis affords exceedingly few "reagents" or tests that may be applied to determine positively that a given cipher belongs to one or the other of two systems yielding externally similar results. And this is what makes the analysis of an isolated, complex cryptogram so difficult. Note the limiting adjective "isolated" in the foregoing sentence, for it is used advisedly. It is not often that the general system fails to disclose itself or cannot be discovered by painstaking investigation when there is a great volume of text accumulating from a regular traffic between numerous correspondents in a large organization. *Sooner or later* the system becomes known, either because of blunders and carelessness on the part of the personnel entrusted with the cryptographing of the messages, or because the accumulation of text itself makes possible the determination of the general system by cryptanalytic studies. But in the case of a single or even a few isolated cryptograms concerning which little or no information can be gained by the cryptanalyst, he is often unable, without a knowledge of, or a shrewd guess as to the general system employed, to decompose the heterogeneous text of the cryptogram into homogeneous, monoalphabetic text, which is the ultimate and essential step in analysis. The only knowledge that the cryptanalyst can bring to his aid in this most difficult step is that gained by long experience and practice in the analysis of many different types of systems.

---

[1] The writer has seen in print statements that "during the World War . . . decoded messages in Japanese and Russian without knowing a word of either language." The extent to which such statements are exaggerated will soon become obvious to the student. Of course, there are occasional instances in which a mere clerk with quite limited experience may be able to "solve" a message in an extremely simple system in a language of which he has no knowledge at all; but such a "solution" calls for nothing more arduous than the ability to recognize pronounceable combinations of vowels and consonants—an ability that hardly deserves to be rated as "cryptanalytic" in any real sense. To say that it is possible to solve a cryptogram in a foreign language "without knowing a word of that language" is not quite the same as to say that it is possible to do so with only a slight knowledge of the language; and it may be stated without cavil that the better the cryptanalyst's knowledge of the language, the greater are the chances for his success and, in any case, the easier is his work.

c. On account of the complexities surrounding this particular phase of cryptanalysis, and because in any scheme of analysis based upon successive eliminations of alternatives the cryptanalyst can only progress so far as the extent of his own knowledge of *all* the possible alternatives will permit, it is necessary that detailed discussion of the eliminative process be postponed until the student has covered most of the field. For example, the student will perhaps want to know at once how he can distinguish between a cryptogram that is in code or enciphered code from one that is in cipher. It is at this stage of his studies impracticable to give him any helpful indications on his question. In return it may be asked of him why he should expect to be able to do this in the early stages of his studies when often the experienced expert cryptanalyst is baffled on the same score!

d. Nevertheless, in lieu of more precise tests not yet discovered, a general guide that may be useful in cryptanalysis will be built up, step by step as the student progresses, in the form of a series of charts comprising what may be designated *An Analytical Key For Cryptanalysis.* (See Par. 50.) It may be of assistance to the student if, as he proceeds, he will carefully study the charts and note the place which the particular cipher he is solving occupies in the general cryptanalytic panorama. These charts admittedly constitute only very brief outlines, and can therefore be of but little direct assistance to him in the analysis of the more complex types of ciphers he may encounter later on. So far as they go, however, they may be found to be quite useful in the study of elementary cryptanalysis. For the experienced cryptanalyst they can serve only as a means of assuring that no possible step or process is inadvertently overlooked in attempts to solve a difficult cipher.

e. Much of the labor involved in cryptanalytic work, as referred to in Par. 2, is connected with this determination of the general system. The preparation of the text, its rewriting in different forms, sometimes being rewritten in a half dozen ways, the recording of letters, the establishment of frequencies of occurrences of letters, comparisons and experiments made with known material of similar character, and so on, constitute much labor that is most often indispensable, but which sometimes turns out to have been wholly unnecessary, or in vain. In a recent treatise [2] it is stated quite boldly that "this work once done, the determination of the system is often relatively easy." This statement can certainly apply only to the simpler types of ciphers; it is entirely misleading as regards the much more frequently encountered complex cryptograms of modern times.

7. **The reconstruction of the specific key.**—*a.* Nearly all practical cryptographic methods require the use of a specific key to guide, control, or modify the various steps under the general system. Once the latter has been disclosed, discovered, or has otherwise come into the possession of the cryptanalyst, the next step in solution is to determine, if necessary, and if possible, the specific key that was employed to cryptograph the message or messages under examination. This determination may not be in complete detail; it may go only so far as to lead to a knowledge of the number of alphabets involved in a substitution cipher, or the number of columns involved in a transposition cipher, or that a one-part code has been used, in the case of a code system. But it is often desirable to determine the specific key in as complete a form and with as much detail as possible, for this information will very frequently be useful in the solution of subsequent cryptograms exchanged between the same correspondents, since the nature of the specific key in a solved case may be expected to give clues to the specific key in an unsolved case.

b. Frequently, however, the reconstruction of the key is not a prerequisite to, and does not constitute an absolutely necessary preliminary step in, the fourth basic operation, *viz,* the reconstruction or establishment of the plain text. In many cases, indeed, the two processes are carried along simultaneously, the one assisting the other, until in the final stages both have been completed in their entireties. In still other cases the reconstruction of the specific key may succeed instead of precede the reconstruction of the plain text, and is accomplished purely as a matter of academic interest; or the specific key may, in unusual cases, never be reconstructed.

8. **The reconstruction of the plain text.**—*a.* Little need be said at this point on this phase of cryptanalysis. The process usually consists, in the case of substitution ciphers, in the establishment of equivalency between specific letters of the cipher text and the plain text, letter by letter, pair by pair, and so on, depending upon the particular type of substitution system involved. In the case of transposition ciphers, the process consists in rearranging the elements of the cipher text, letter by letter, pair by pair, or occasionally word by word, depending upon the particular type of transposition system involved, until the letters or words have been returned to their original plain-text order. In the case of code, the process consists in determining the meaning of each code group and inserting this meaning in the code text to reestablish the original plain text.

b. The foregoing processes do not, as a rule, begin at the beginning of a message and continue letter by letter, or group by group in sequence up to the very end of the message. The establishment of values of cipher letters in substitution methods, or of the positions to which cipher letters should be transferred to form the plain text in the case of transposition methods, comes at very irregular intervals in the process. At first only one or two values scattered here and there throughout the text may appear; these then form the "skeletons" of words, upon which further work, by a continuation of the reconstruction process, is made possible; in the end the complete or nearly complete [3] text is established.

c. In the case of cryptograms in a foreign language, the translation of the solved messages is a final and necessary step, but is not to be considered as a cryptanalytic process. However, it is commonly the case that the translation process will be carried on simultaneously with the cryptanalytic, and will aid the latter, especially when there are lacunae which may be filled in from the context. (See also Par. 5c in this connection.)

---

[2] Lange et Soudart, *op. cit.*, p. 106.

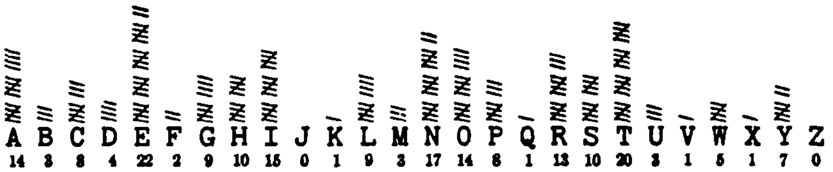
[3] Sometimes in the case of code, the meaning of a few code groups may be lacking, because there is insufficient text to establish their meaning.

SECTION III

FREQUENCY DISTRIBUTIONS

Paragraph
The simple or uniliteral frequency distribution 9
Important features of the normal uniliteral frequency distribution 10
Constancy of the standard or normal uniliteral frequency distribution 11

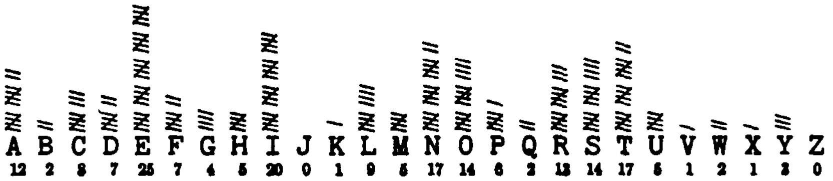
9. The simple or uniliteral frequency distribution.—*a*. It has long been known to cryptographers and typographers that the letters composing the words of any intelligible written text composed in any language which is alphabetic in construction are employed with greatly varying frequencies. For example, if on cross-section paper a simple tabulation, shown in Fig. 1, called a *uniliteral frequency distribution*, is made of the letters composing the words of the preceding sentence, the variation in frequency is strikingly demonstrated. It is seen that whereas certain letters, such as A, E, I, N, O, R, S, and T, are employed very frequently, other letters, such as C, G, P, and W are employed not nearly so frequently, while still other letters, such as F, J, Q, V, and Z are employed either seldom or not at all.



(Total=200 letters)

FIGURE 1.

b. If a similar tabulation is now made of the letters comprising the words of the second sentence in the preceding paragraph, the graph shown in Fig. 2 is obtained. Both sentences have exactly the same number of letters (200).



(Total=200 letters)

FIGURE 2.

c. Although each of these two graphs exhibits great variation in the relative frequencies with which *different* letters are employed in the respective sentences to which they apply, no marked differences are exhibited between the frequencies of the *same* letter in the two graphs. Compare, for example, the frequencies of A, B, C . . . Z in Fig. 1 with those of A, B, C, . . . Z in Fig. 2. Aside from one or two exceptions, as in the case of the letter F, these two graphs agree rather strikingly.

d. This agreement, or *similarity*, would be practically complete if the two texts were much longer, for example, five times as long. In fact, when two texts of similar character, each containing more than 1,000 letters, are compared, it would be found that the respective frequencies of the 26 letters composing the two graphs show only very slight differences. This means, in other words, that in normal, plain text each letter of the alphabet occurs with a rather *constant* or *characteristic frequency* which it tends to approximate, depending upon the length of the text analyzed. The longer the text (within certain limits), the closer will be the approximation to the characteristic frequencies of letters in the language involved. However, when the amount of text being analyzed has reached a substantial volume (roughly, 1,000 letters), the practical gain in accuracy does not warrant further increase in the amount of text.[1]

e. An experiment along these lines will be convincing. A series of 260 official telegrams[2] passing through the War Department Message Center was examined statistically. The messages were divided into five sets, each totaling 10,000 letters, and the five distributions shown in Table 1–A, were obtained.

f. If the five distributions in Table 1–A are summed, the results are as shown in Table 2–A.

TABLE 1–A.—*Absolute frequencies of letters appearing in five sets of Governmental plain-text telegrams, each set containing 10,000 letters, arranged alphabetically*

Message No. 1		Message No. 2		Message No. 3		Message No. 4		Message No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1,367	E	1,294	E	1,292	E	1,270	E	1,275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	233	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10,000		10,000		10,000		10,000		10,000

---

[1] See footnote 5, page 16.

[2] These comprised messages from several departments in addition to the War Department and were all of an administrative character.

TABLE 2-A.—*Absolute frequencies of letters appearing in the combined five sets of messages totaling 50,000 letters, arranged alphabetically*

A	3,683	G	819	L	1,821	Q	175	V	766
B	487	H	1,694	M	1,237	R	3,788	W	780
C	1,534	I	3,676	N	3,975	S	3,058	X	231
D	2,122	J	82	O	3,764	T	4,595	Y	967
E	6,498	K	148	P	1,335	U	1,300	Z	49
F	1,416								

g. The frequencies noted in subparagraph *f*, when reduced to the basis of 1,000 letters and then used as a basis for constructing a simple chart that will exhibit the variations in frequency in a striking manner, yield the following graph which is hereafter designated as the *normal*, or *standard uniliteral frequency distribution* for English telegraphic plain text:

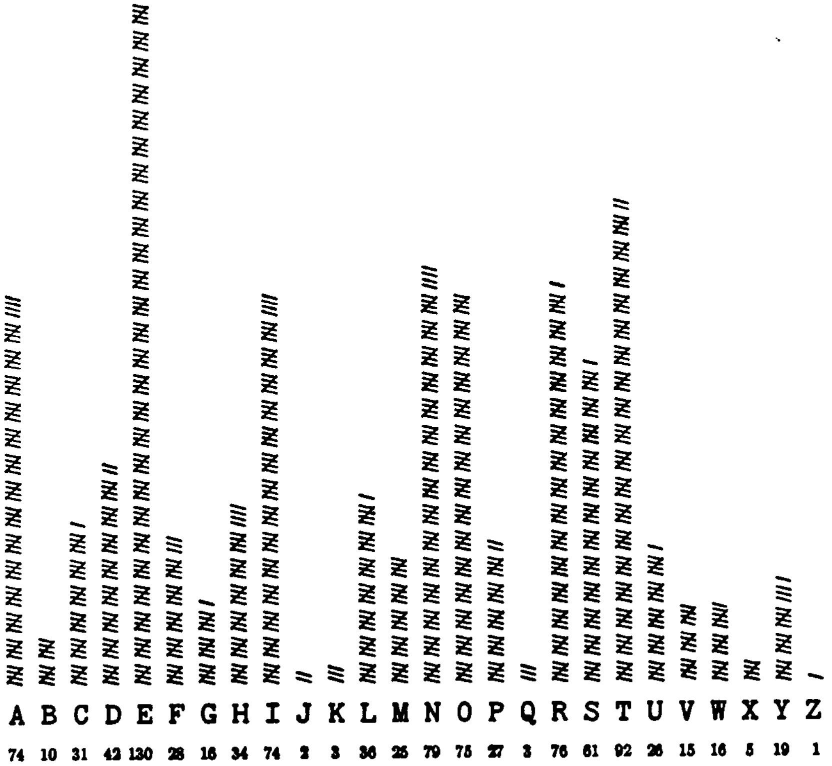


FIGURE 3.

10. Important features of the normal uniliteral frequency distribution.—*a.* When the graph shown in Fig. 3 is studied in detail, the following features are apparent:

(1) It is quite irregular in appearance. This is because the letters are used with greatly varying frequencies, as discussed in the preceding paragraph. This irregular appearance is often described by saying that the graph shows marked *crests and troughs*, that is, points of high frequency and low frequency.

(2) The relative positions in which the crests and troughs fall within the graph, that is, the *spatial relations* of the crests and troughs, are rather definitely fixed and are determined by circumstances which have been explained in a preceding text.[3]

(3) The relative heights and depths of the crests and troughs within the graph, that is, the *linear extensions* of the lines marking the respective frequencies, are also rather definitely fixed, as would be found if an equal volume of similar text were analyzed.

(4) The most prominent crests are marked by the vowels A, E, I, O, and the consonants N, R, S, T; the most prominent troughs are marked by the consonants J, K, Q, X, and Z.

(5) The important data are summarized in tabular form in Table 3.

TABLE 3

	Frequency	Percent of total	Percent of total in round numbers
6 Vowels: A E I O U Y	398	39.8	40
20 Consonants:			
5 High Frequency (D N R S T)	350	35.0	35
10 Medium Frequency (B C F G H L M P V W)	238	23.8	24
5 Low Frequency (J K Q X Z)	14	1.4	1
Total	1,000	100.0	100

(6) The frequencies of the letters of the alphabet are as follows:

A	74	G	16	L	36	Q	3	V	15
B	10	H	34	M	25	R	76	W	16
C	31	I	74	N	79	S	61	X	5
D	42	J	2	O	75	T	92	Y	19
E	130	K	3	P	27	U	26	Z	1
F	28								

(7) The relative order of frequency of the letters is as follows:

E	130	I	74	C	31	Y	19	X	5
T	92	S	61	F	28	G	16	Q	3
N	79	D	42	P	27	W	16	K	3
R	76	L	36	U	26	V	15	J	2
O	75	H	34	M	25	B	10	Z	1
A	74								

(8) The four vowels A, E, I, O (combined frequency 353) and the four consonants N, R, S, T (combined frequency 308) form 661 out of every 1,000 letters of plain text; in other words, *less than ⅓ of the alphabet is employed in writing ⅔ of normal plain text.*

---

[3] Section VII, *Elementary Military Cryptography.*

b. The data given in Fig. 3 and Table 3 represent the relative frequencies found in a large volume of English telegraphic text of a governmental, administrative character. These frequencies will vary somewhat with the nature of the text analyzed. For example, if an equal number of telegrams dealing solely with *commercial* transactions in the *leather industry* were studied statistically, the frequencies would be slightly different because of the repeated occurrence of words peculiar to that industry. Again, if an equal number of telegrams dealing solely with *military* messages of a *tactical* character were studied statistically, the frequencies would differ slightly from those found above for general governmental messages of an administrative character.

c. If ordinary English literary text (such as may be found in any book, newspaper, or printed document) were analyzed, the frequencies of certain letters would be changed to an appreciable degree. This is because in telegraphic text words which are not strictly essential for intelligibility (such as the definite and indefinite articles, certain prepositions, conjunctions and pronouns) are omitted. In addition, certain essential words, such as "stop", "period", "comma", and the like, which are usually indicated in written or printed matter by symbols not easy to transmit telegraphically and which must, therefore, be spelled out in telegrams, occur very frequently. Furthermore, telegraphic text often employs longer and more uncommon words than does ordinary newspaper or book text.

d. As a matter of fact, other tables compiled in the Office of the Chief Signal Officer gave slightly different results, depending upon the source of the text. For example, three tables based upon 75,000, 100,000, and 136,257 letters taken from various sources (telegrams, newspapers, magazine articles, books of fiction) gave as the relative order of frequency for the first 10 letters the following:

For 75,000 letters	E T R N I O A S D L
For 100,000 letters	E T R I N O A S D L
For 136,257 letters	E T R N A O I S L D

TABLE 4.—*Frequency table for 10,000 letters of literary English, as compiled by Hitt*

ALPHABETICALLY ARRANGED

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

Hitt also compiled data for telegraphic text (but does not state what kind of messages) and gives the following table:

TABLE 5.—*Frequency table for 10,000 letters of telegraphic English, as compiled by Hitt*

ALPHABETICALLY ARRANGED

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

e. Frequency data applicable purely to English military text were compiled by Hitt,[4] from a study of 10,000 letters taken from orders and reports. The frequencies found by him are given in Tables 4 and 5.

11. Constancy of the standard or normal, uniliteral frequency distribution.—*a.* The relative frequencies disclosed by the statistical study of large volumes of text may be considered to be the standard or *normal* frequencies of the letters of written English. Counts made of smaller volumes of text will tend to approximate these normal frequencies, and, within certain limits,[5] the smaller the volume, the lower will be the degree of approximation to the normal, until, in the case of a very short message, the normal proportions may not obtain at all. It is advisable that the student fix this fact firmly in mind, for the sooner he realizes the true nature of any data relative to the frequency of occurrence of letters in text, the less often will his labors toward the solution of specific ciphers be thwarted and retarded by too strict an adherence to these generalized principles of frequency. He should constantly bear in mind that such data are merely statistical generalizations, that they will be found to hold strictly true only in large volumes of text, and that they may not even be approximated in short messages.

b. Nevertheless the normal frequency distribution or the "normal expectancy" for any alphabetic language is, in the last analysis, the best guide to, and the usual basis for, the solution of cryptograms of a certain type. It is useful, therefore, to reduce the normal, uniliteral frequency distribution to a basis that more or less closely approximates the volume of text which the cryptanalyst most often encounters in individual cryptograms. As regards length of messages, counting only the letters in the body, and excluding address and signature, a study of the

---

[4] *Op. cit.*, pp. 6–7.

[5] It is useless to go beyond a certain limit in establishing the normal-frequency distribution for a given language. As a striking instance of this fact, witness the frequency study made by an indefatigable German, Kaeding, who in 1898 made a count of the letters in about 11,000,000 words, totaling about 62,000,000 letters in German text. When reduced to a percentage basis, and when the relative order of frequency was determined, the results he obtained differed very little from the results obtained by Kasiski, a German cryptographer, from a count of only 1,060 letters. See Kaeding, *Haeufigkeitswoerterbuch*, Steglitz, 1898; Kasiski, *Die Geheimschriften und die Dechiffrir-Kunst*, Berlin, 1863.

260 telegrams referred to in paragraph 9 shows that the arithmetical average is 217 letters; the statistical mean, or weighted average,[6] however, is 191 letters. These two results are, however, close enough together to warrant the statement that the *average* length of telegrams is approximately 200 letters. The frequencies given in Par. 9*f* have therefore been reduced to a basis of 200 letters, and the following uniliteral frequency distribution may be taken as showing the most typical distribution to be expected in 200 letters of telegraphic English text:

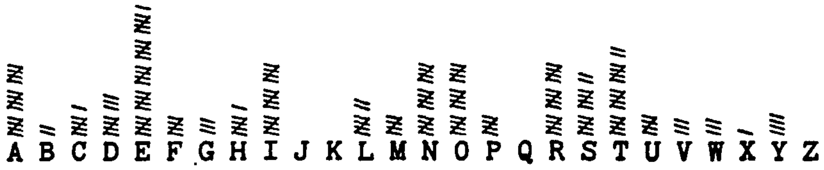


FIGURE 4.

c. The student should take careful note of the appearance of the distribution[7] shown in Fig. 4, for it will be of much assistance to him in the early stages of his study. The manner of setting down the tallies should be followed by him in making his own distributions, indicating every fifth occurrence of a letter by an oblique tally. This procedure almost automatically shows the total number of occurrences for each letter, and yet does not destroy the graphical appearance of the distribution, especially if care is taken to use approximately the same amount of space for each set of five tallies. Cross-section paper is very useful for this purpose.

d. The word "uniliteral" in the designation "uniliteral frequency distribution" means "single letter", and it is to be inferred that other types of frequency distributions may be encountered. For example, a distribution of pairs of letters, constituting a biliteral frequency distribution, is very often used in the study of certain cryptograms in which it is desired that pairs made by combining successive letters be listed. A biliteral distribution of A B C D E F would take these pairs: AB, BC, CD, DE, EF. The distribution could be made in the form of a large square divided up into 676 cells. When distributions beyond biliteral are required (triliteral, quadraliteral, etc.) they can only be made by listing them in some order, for example, alphabetically based on the 1st, 2d, 3d, . . . letter.

---

[6] The arithmetical average is obtained by adding each different length and dividing by the number of different-length messages; the mean is obtained by multiplying each different length by the number of messages of that length, adding all products, and dividing by the total number of messages.

[7] The use of the terms "distribution" and "frequency distribution", instead of "table" and "frequency table", respectively, is considered advisable from the point of view of consistency with the usual statistical nomenclature. When data are given in tabular form, with frequencies indicated by numbers, then they may properly be said to be set out in the form of a *table*. When, however, the same data are distributed in a chart which partakes of the nature of a graph, with the data indicated by horizontal or vertical linear extensions, or by a curve connecting points corresponding to quantities, then it is more proper to call such a graphic representation of the data a *distribution*.

SECTION IV

FUNDAMENTAL USES OF THE UNILITERAL FREQUENCY DISTRIBUTION

Paragraph

The four facts which can be determined from a study of the uniliteral frequency distribution for a cryptogram 12

Determining the class to which a cipher belongs 13

Determining whether a substitution cipher is monoalphabetic or polyalphabetic 14

Determining whether the cipher alphabet is a standard, or a mixed cipher alphabet 15

Determining whether the standard cipher alphabet is direct or reversed 16

12. The four facts which can be determined from a study of the uniliteral frequency distribution for a cryptogram. *a.* The following four facts (to be explained subsequently) can usually be determined from an inspection of the uniliteral frequency distribution for a given cipher message of average length, composed of letters:

(1) Whether the cipher belongs to the substitution or the transposition class;

(2) If to the former, whether it is monoalphabetic or polyalphabetic in character;

(3) If monoalphabetic, whether the cipher alphabet is a standard cipher alphabet or a mixed cipher alphabet;

(4) If standard, whether it is a direct or reversed standard cipher alphabet.

b. For immediate purposes the first two of the foregoing determinations are quite important and will be discussed in detail in the next two subparagraphs; the other two determinations will be touched upon very briefly, leaving their detailed discussion for subsequent sections of the text.

13. Determining the class to which a cipher belongs.—*a.* The determination of the class to which a cipher belongs is usually a relatively easy matter because of the fundamental difference in the nature of transposition and of substitution as cryptographic processes. In a transposition cipher the original letters of the plain text have merely been rearranged, without any change whatsoever in their identities, that is, in the conventional values they have in the normal alphabet. Hence, the numbers of vowels (A, E, I, O, U, Y), high-frequency consonants (D, N, R, S, T), medium-frequency consonants (B, C, F, G, H, L, M, P, V, W), and low-frequency consonants (J, K, Q, X, Z) are exactly the same in the cryptogram as they are in the plain-text message. Therefore, the percentages of vowels, high, medium, and low-frequency consonants are the same in the transposed text as in the equivalent plain text. In a substitution cipher, on the other hand, the identities of the original letters of the plain text have been changed, that is, the conventional values they have in the normal alphabet have been altered. Consequently, if a count is made of the various letters present in such a cryptogram, it will be found that the number of vowels, high, medium, and low-frequency consonants will usually be quite different in the cryptogram from what they are in the original plain-text message. Therefore, the percentages of vowels, high, medium, and low-frequency consonants are usually quite different in the substitution text from what they are in the equivalent plain text. From these considerations it follows that if in a specific cryptogram the percentages of vowels, high, medium, and low-frequency consonants are approximately the same as would be expected in normal plain text, the cryptogram *probably* belongs to the transposition class; if these percentages are quite different from those to be expected in normal plain text the cryptogram *probably* belongs to the substitution class.

b. In the preceding subparagraph the word "probably" was emphasized by italicizing it, for there can be no certainty in every case of this determination. *Usually* these percentages in a transposition cipher are close to the normal percentages for plain text; *usually*, in a substitution cipher, they are far different from the normal percentages for plain text. But occasionally a cipher message is encountered which is difficult to classify with a reasonable degree of certainty because the message is too short for the general principles of frequency to manifest themselves. It is clear that if in actual messages there were no variation whatever from the normal vowel and consonant percentages given in Table 3, the determination of the class to which a specific cryptogram belongs would be an extremely simple matter. But unfortunately there is always some variation or deviation from the normal. Intuition suggests that as messages decrease in length there may be a greater and greater departure from the normal proportions of vowels, high, medium, and low-frequency consonants, until in very short messages the normal proportions may not hold at all. Similarly, as messages increase in length there may be a lesser and lesser departure from the normal proportions, until in messages totalling a thousand or more letters there may be no difference at all between the actual and the theoretical proportions. But intuition is not enough, for in dealing with specific messages of the length of those commonly encountered in practical work the question sometimes arises as to exactly how much deviation (from the normal proportions) may be allowed for in a cryptogram which shows a considerable amount of deviation from the normal and which might still belong to the transposition rather than to the substitution class.

c. Statistical studies have been made on this matter and some graphs have been constructed thereon. These are shown in Charts 1–4 in the form of simple curves, the use of which will now be explained. Each chart contains two curves marking the lower and upper limits, respectively, of the theoretical amount of deviation (from the normal percentages) of vowels or consonants which may be allowable in a cipher believed to belong to the transposition class.

d. In Chart 1, curve V_1 marks the lower limit of the theoretical amount of deviation from the normal number of vowels to be expected in a message of given length; curve V_2 marks the upper limit of the same thing. Thus, for example, in a message of 100 letters in plain English there should be between 33 and 47 vowels (A E I O U Y). Likewise, in Chart 2 curves H_1 and H_2 mark the lower and upper limits as regards the high-frequency consonants. In a message of 100 letters there should be between 28 and 42 high-frequency consonants (D N R S T). In Chart 3, curves M_1 and M_2 mark the lower and upper limits as regards the medium-frequency consonants. In a message of 100 letters there should be between 17 and 31 medium-frequency consonants (B C F G H L M P V W). Finally, in Chart 4, curves L_1 and L_2 mark the lower and upper limits as regards the low-frequency consonants. In a message of 100 letters there should be between 0 and 3 low-frequency consonants (J K Q X Z). In using the charts, therefore, one finds the point of intersection of the vertical coordinate corresponding to the length of the message, with the horizontal coordinate corresponding to (1) the number of vowels, (2) the number of high-frequency consonants, (3) the number of medium-frequency consonants, and (4) the number of low-frequency consonants actually counted in the message. If all four points of intersection fall within the area delimited by the respective curves, then the number of vowels, high, medium, and low-frequency consonants corresponds with the number theoretically expected in a normal plain-text message of the same length; since the message under investigation is not plain text, it follows that the cryptogram may certainly be classified as a transposition cipher. On the other hand, if one or more of these points of intersection falls outside the area delimited by the respective curves, it follows that the cryptogram is probably a substitution cipher. The distance that the point of intersection falls outside the area delimited by these curves is a more or less rough measure of the improbability of the cryptogram's being a transposition cipher.

e. Sometimes a cryptogram is encountered which is hard to classify with certainty even with the foregoing aids, because it has been consciously prepared with a view to making the classification difficult. This can be done either by selecting peculiar words (as in "trick cryptograms") or by employing a cipher alphabet in which letters of *approximately similar normal frequencies* have been interchanged. For example, E may be replaced by O, T by R, and so on, thus yielding a cryptogram giving external indications of being a transposition cipher but which is really a substitution cipher. If the cryptogram is not too short, a close study will usually disclose what has been done, as well as the futility of so simple a subterfuge.

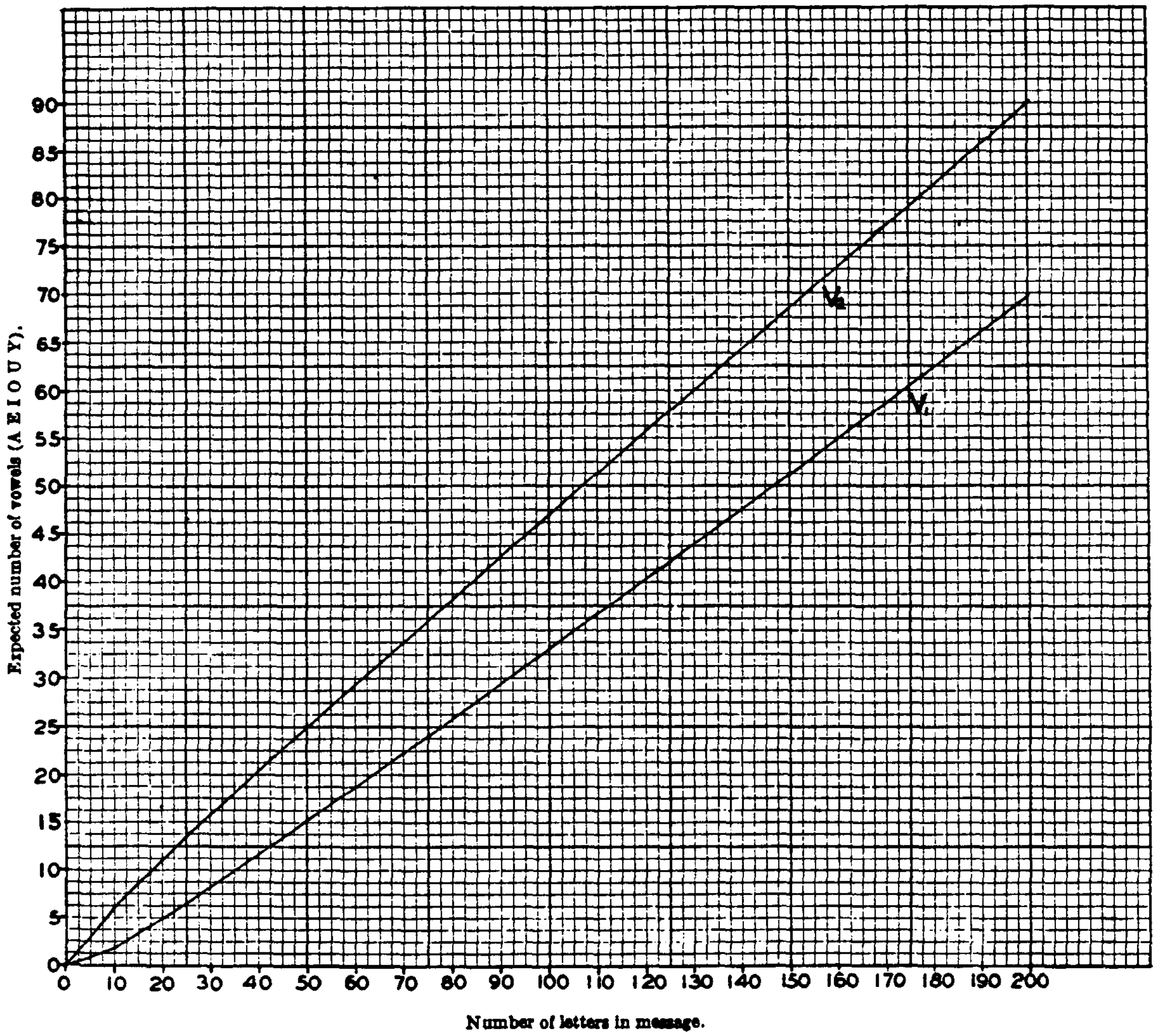


CHART No. 1.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of vowels to be expected in messages of various lengths. (See Par. 13*d*.)

f. In the majority of cases, in practical work, the determination of the class to which a cipher of average length belongs can be made from a mere inspection of the message, after the cryptanalyst has acquired a familiarity with the normal appearance of transposition and of substitution ciphers. In the former case, his eyes very speedily note many high-frequency letters, such as E, T, N, R, O, and S, with the absence of low-frequency letters, such as J, K, Q, X, and Z; in the latter case, his eyes just as quickly note the presence of many low-frequency letters, and a corresponding absence of the usual high-frequency letters.

g. Another rather quickly completed test, in the case of the simpler varieties of ciphers, is to look for *repetitions of groups of letters*. As will become apparent very soon, recurrences of syllables, entire words and short phrases constitute a characteristic of all normal plain text. Since a transposition cipher involves a change in the *sequence* of the letters composing a plaintext message, such recurrences are broken up so that the cipher text no longer will show repetitions of more or less lengthy sequences of letters. But if a cipher message does show many repetitions and these are of several letters in length, say over four or five, the conclusion is at once warranted that the cryptogram is most probably a substitution and not a transposition cipher. However, for the beginner in cryptanalysis, it will be advisable to make the uniliteral frequency distribution, and note the frequencies of the vowels, the high, medium, and low-frequency consonants. Then, referring to Charts 1 to 4, he should carefully note whether or not the observed frequencies for

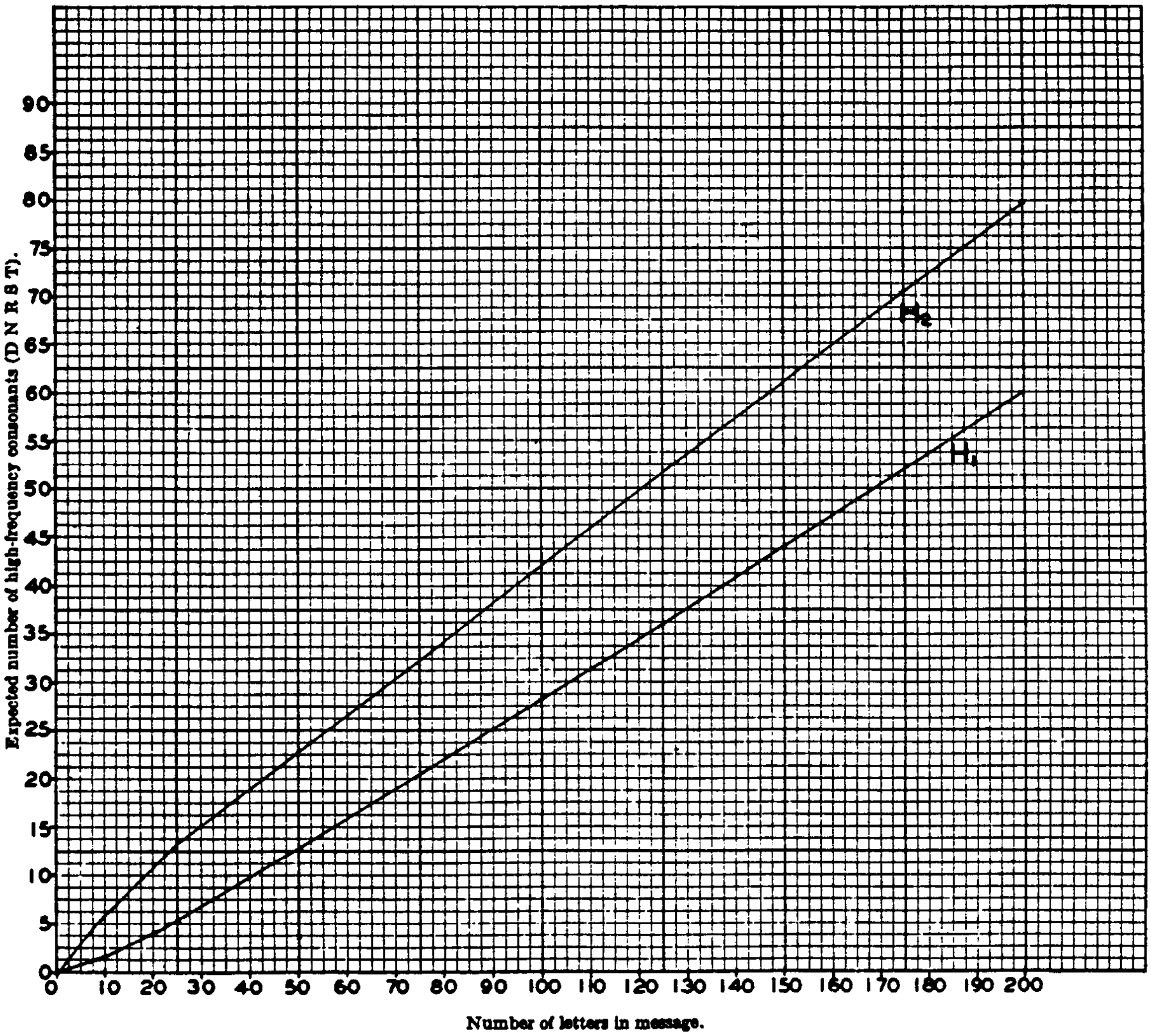


CHART No. 2.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of high-frequency consonants to be expected in messages of various lengths. (See Par. 13*d*.)

these categories of letters fall within the limits of the theoretical frequencies for a normal plain-text message of the same length, and be guided accordingly.

h. It is obvious that the foregoing rule applies only to ciphers composed wholly of letters. If a message is composed entirely of figures, or of arbitrary signs and symbols, or of intermixtures of letters, figures and other symbols, it is immediately apparent that the cryptogram is a substitution cipher.

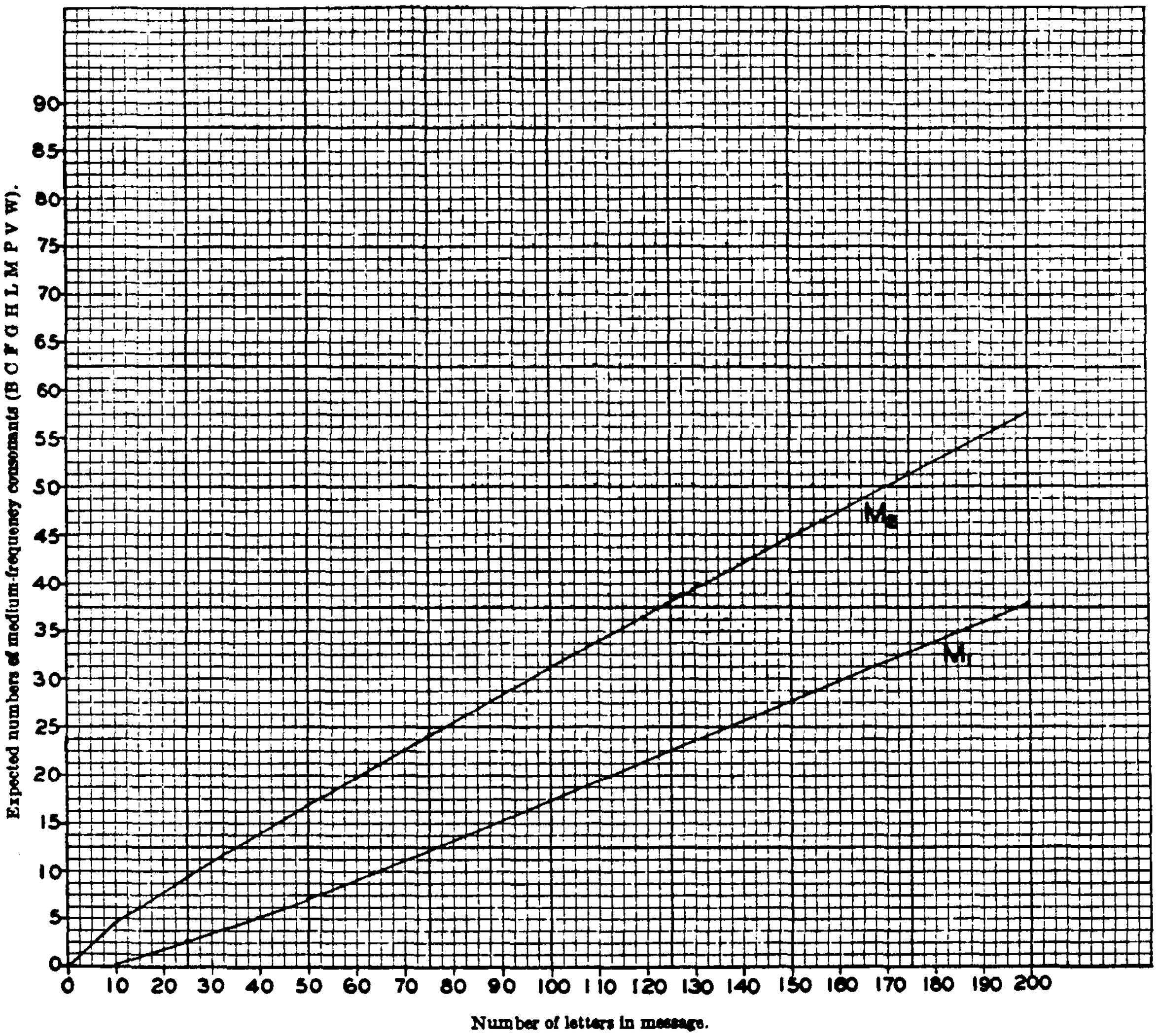


CHART No. 3.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of medium-frequency consonants to be expected in messages of various lengths. (See Par. 13*d*.)

i. Finally, it should be mentioned that there are certain kinds of cryptograms whose class cannot be determined by the method set forth in subparagraphs *b, c, d* above. These exceptions will be discussed in a subsequent section of this text.[1]

14. Determining whether a substitution cipher is monoalphabetic or polyalphabetic.—*a.* It will be remembered that a monoalphabetic substitution cipher is one in which a single cipher alphabet is employed throughout the whole message, that is, a given plain-text letter is invariably represented throughout the message by one and the same letter in the cipher text. On the other hand, a polyalphabetic substitution cipher is one in which two or more cipher alphabets are employed within the same message; that is, a given plain-text letter may be represented by two or more different letters in the cipher text, according to some rule governing the selection of the equivalent to be used in each case. From this it follows that a single cipher letter may represent two or more different plain-text letters.

---

[1] Par. 47.

b. It is easy to see why and how the appearance of the uniliteral frequency distribution for a substitution cipher may be used to determine whether the cryptogram is monoalphabetic or polyalphabetic in character. The normal distribution presents marked crests and troughs by virtue of two circumstances. First, the elementary sounds which the symbols represent are used with greatly varying frequencies, it being one of the striking characteristics of every alphabetic language that its elementary sounds are used with greatly varying frequencies.[2] In the second place, except for orthographic aberrations peculiar to certain languages (conspicuously, English and French), each such sound is represented by the same symbol. It follows, therefore, that since in a monoalphabetic substitution cipher each different plain-text letter (=elementary sound) is represented by one and only one cipher letter (=elementary symbol), the uniliteral frequency distribution for such a cipher message must also exhibit the irregular crest and trough appearance of the normal distribution, but with only this important modification—*the absolute*

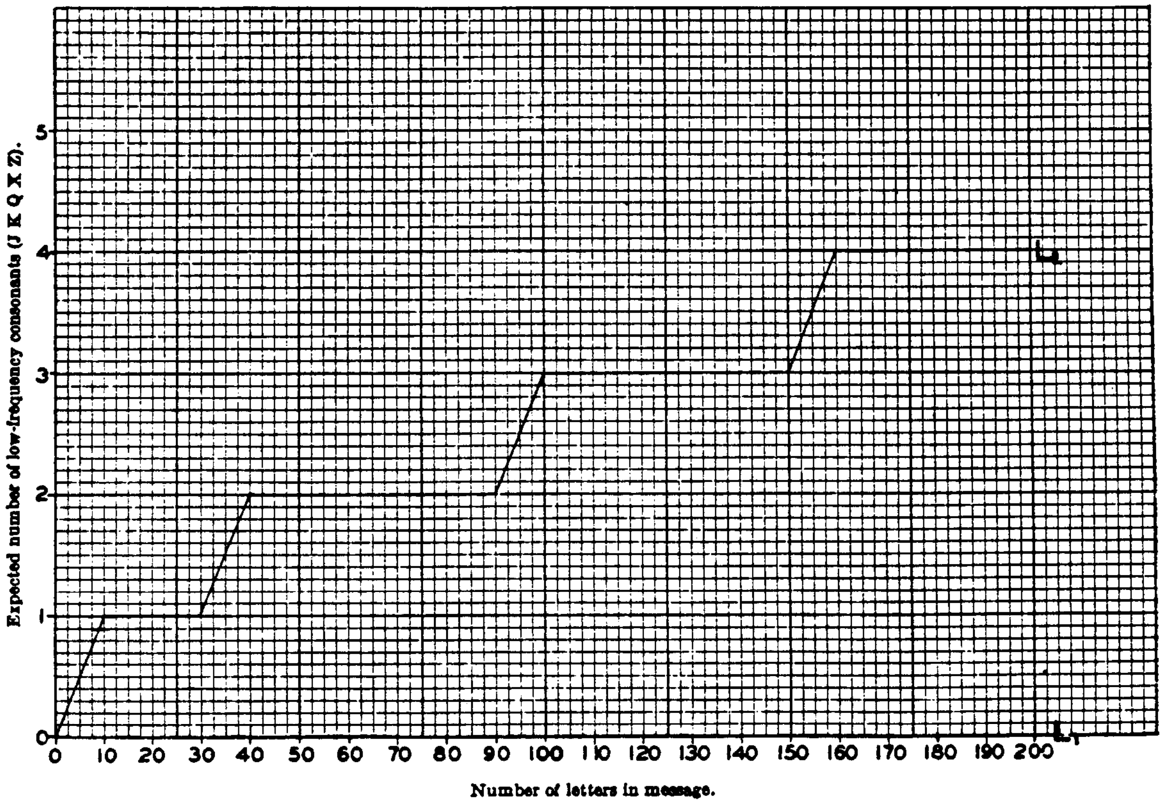


CHART No. 4.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of low-frequency consonants to be expected in messages of various lengths. (See Par. 13*d*.)

---

[2] The student who is interested in this phase of the subject may find the following reference of value: Zipf' G. K., *Selected Studies of the Principle of Relative Frequency in Language*, Cambridge, Mass., 1932.

positions of the crests and troughs will not be the same as in the normal. That is, the letters accompanying the crests and the troughs in the distribution for the cryptogram will be different from those accompanying the crests and the troughs in the normal distribution. But the marked irregularity of the distribution, the presence of accentuated crests and troughs, is in itself an indication that each symbol or cipher letter always represents the same plain-text letter in that cryptogram. Hence the general rule: *A marked crest and trough appearance in the uniliteral frequency distribution for a given cryptogram indicates that a single cipher alphabet is involved and constitutes one of the tests for a monoalphabetic substitution cipher.*

c. On the other hand, suppose that in a cryptogram each cipher letter represents several different plain-text letters. Some of them are of high frequency, others of low frequency. The net result of such a situation, so far as the uniliteral frequency distribution for the cryptogram is concerned, is to prevent the appearance of any marked crests and troughs and to tend to reduce the elements of the distribution to a more or less common level. This imparts a "flattened out" appearance to the distribution. For example, in a certain cryptogram of polyalphabetic construction, $K_c=E_p$, G_p, and J_p; $R_c=A_p$, D_p, and B_p; $X_c=O_p$, L_p, and F_p. The frequencies of K_c, R_c, and X_c will be approximately equal because the summations of the frequencies of the several plain-text letters each of these cipher letters represents at different times will be about equal. If this same phenomenon were true of all the letters of the cryptogram, it is clear that the frequencies of the 26 letters, when shown by means of the ordinary uniliteral frequency distribution, would show no striking differences and the distribution would have the flat appearance of a typical polyalphabetic substitution cipher. Hence, the general rule: *The absence of marked crests and troughs in the uniliteral frequency distribution indicates that two or more cipher alphabets are involved. The flattened-out appearance of the distribution constitutes one of the tests for a polyalphabetic substitution cipher.*

d. The foregoing test based upon the appearance of the frequency distribution constitutes only one of several means of determining whether a substitution cipher is monoalphabetic or polyalphabetic in composition. It can be employed in cases yielding frequency distributions from which definite conclusions can be drawn with more or less certainty by mere ocular examination. In those cases in which the frequency distributions contain insufficient data to permit drawing definite conclusions by such examination, certain statistical tests can be applied. These will be discussed in a subsequent text.

e. At this point, however, one additional test will be given because of its simplicity of application. It may be employed in testing messages up to 200 letters in length, it being assumed that in messages of greater length ocular examination of the frequency distribution offers little or no difficulty. This test concerns the *number of blanks* in the frequency distribution, that is, the number of letters of the alphabet which are entirely absent from the message. It has been found from statistical studies that rather definite "laws" govern the theoretically expected number of blanks in normal plain-text messages and in frequency distributions for cryptograms of different natures and of various sizes. The results of certain of these studies have been embodied in Chart 5.

f. This chart contains two curves. The one labeled *P* applies to the average number of blanks theoretically expected in frequency distributions based upon normal plain-text messages of the indicated lengths. The other curve, labeled *R*, applies to the average number of blanks theoretically expected in frequency distributions based upon perfectly *random* assortments of letters; that is, assortments such as would be found by random selection of letters out of a hat containing thousands of letters, all of the 26 letters of the alphabet being present in equal proportions, each letter being replaced after a record of its selection has been made. Such random assortments correspond to polyalphabetic cipher messages in which the number of cipher alphabets is so large that if uniliteral frequency distributions are made of the letters, the distributions are practically identical with those which are obtained by random selections of letters out of a hat.

g. In using this chart, one finds the point of intersection of the vertical coordinate corresponding to the length of the message, with the horizontal coordinate corresponding to the observed number of blanks in the distribution for the message. If this point of intersection falls closer to curve *P* than it does to curve *R*, the number of blanks in the message approximates or corresponds more closely to the number theoretically expected in a plain-text message than it does to a random (cipher-text) message of the same length; therefore, this is evidence that the cryptogram is monoalphabetic. Conversely, if this point of intersection falls closer to curve *R* than to curve *P*, the number of blanks in the message approximates or corresponds more closely to the number theoretically expected in a random text than it does to a plain-text message of the same length; therefore, this is evidence that the cryptogram is polyalphabetic.

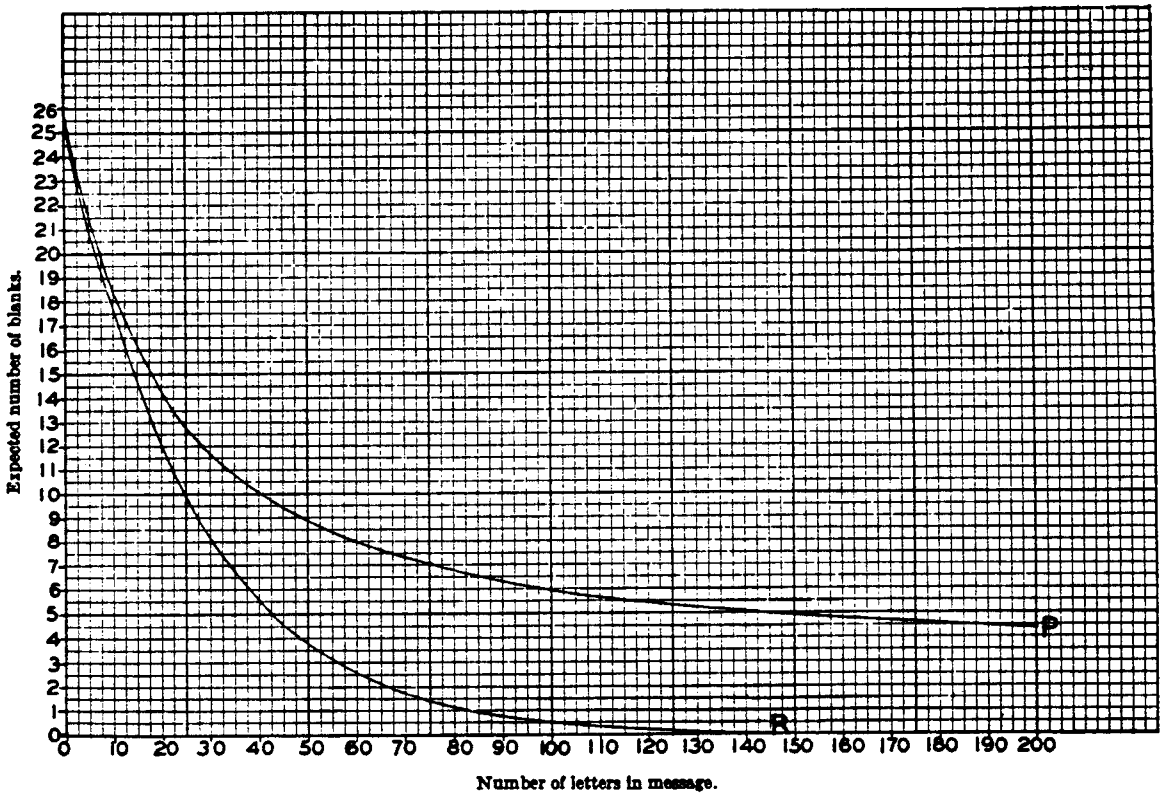


CHART No. 5.—Curves showing the average number of blanks theoretically expected in distributions for plain text (P) and for random text (R) for messages of various lengths. (See Par. 14*f*.)

h. Practical examples of the use of this chart will be given in some of the illustrative messages to follow.

15. **Determining whether the cipher alphabet is a standard, or a mixed cipher alphabet.**—*a*. Assuming that the uniliteral frequency distribution for a given cryptogram has been made, and that it shows clearly that the cryptogram is a substitution cipher and is monoalphabetic in character, a consideration of the nature of standard cipher alphabets[3] almost makes it obvious how an inspection of the distribution will disclose whether the cipher alphabet involved is a standard cipher alphabet or a mixed cipher alphabet. If the crests and troughs of the distribu-

---

[3] See Sec. VIII, *Elementary Military Cryptography*.

tion occupy positions which correspond to the *relative* positions they occupy in the normal frequency distribution, then the cipher alphabet is a standard cipher alphabet. If this is not the case, then it is highly probable that the cryptogram has been prepared by the use of a mixed cipher alphabet.

b. A mechanical test may be applied in doubtful cases arising from lack of material available for study. Just what this test involves, and an illustration of its application will be given in the next section, using specific examples.

16. Determining whether the standard cipher alphabet is direct or reversed.—Assuming that the frequency distribution for a given cryptogram shows clearly that a standard cipher alphabet is involved, the determination as to whether the alphabet is direct or reversed can also be made by inspection, since the difference between the two is merely a matter of the *direction* in which the sequence of crests and troughs progresses—to the right, as in normal reading or writing, or the left. In a direct cipher alphabet the direction in which the crests and troughs of the distribution should be read is the normal direction, from left to right; in a reversed cipher alphabet this direction is reversed, from right to left.

SECTION V

UNILITERAL SUBSTITUTION WITH STANDARD CIPHER ALPHABETS

Paragraph
Principles of solution by construction and analysis of the uniliteral frequency distribution 17
Theoretical example of solution 18
Practical example of solution by the frequency method 19
Solution by completing the plain-component sequence 20
Special remarks on the method of solution by completing the plain-component sequence 21
Value of mechanical solution as a short cut 22

17. Principles of solution by construction and analysis of the uniliteral frequency distribution.—*a*. Standard cipher alphabets are of two sorts, direct and reversed. The analysis of monoalphabetic cryptograms prepared by their use follows almost directly from a consideration of the nature of such alphabets. Since the cipher component of a standard cipher alphabet consists either of the normal sequence merely displaced 1, 2, 3, . . . intervals from the normal point of coincidence, or of the normal sequence proceeding in a reversed-normal direction, it is obvious that the uniliteral frequency distribution for a cryptogram prepared by means of such a cipher alphabet employed monoalphabetically will show crests and troughs whose *relative* positions and frequencies will be exactly the same as in the uniliteral frequency distribution for the plain text of that cryptogram. The only thing that has happened is that the whole set of crests and troughs of the distribution has been displaced to the right or left of the position it occupies in the distribution for the plain text; or else the successive elements of the whole set progress in the opposite direction. Hence, it follows that the correct determination of the plain-text value of the letter marking *any* crest or trough of the uniliteral frequency distribution will result at one stroke in the correct determination of the plain-text values of *all* the remaining 25 letters respectively marking the other crests and troughs in that distribution. Thus, having determined the value of a single element of the cipher component of the cipher alphabet, the values of all the remaining letters of the cipher component are automatically solved at one stroke. In more simple language, the correct determination of the value of a single letter of the cipher text automatically gives the values of the other 25 letters of the cipher text. The problem thus resolves itself into a matter of selecting that point of attack which will most quickly or most easily lead to the determination of the value of *one* cipher letter. The single word *identification* will hereafter be used for the phrase "determination of the value of a cipher letter"; to *identify* a cipher letter is to find its plain-text value.

b. It is obvious that the easiest point of attack is to assume that the letter marking the crest of greatest frequency in the frequency distribution for the cryptogram represents E_p. Proceeding from this initial point, the identifications of the remaining cipher letters marking the other crests and troughs are tentatively made on the basis that the letters of the cipher component proceed in accordance with the normal alphabetic sequence, either direct or reversed. If the actual frequency of each letter marking a crest or a trough approximates to a fairly close degree the normal theoretical frequency of the assumed plain-text equivalent, then the initial identification $\theta_c = E_p$ may be *assumed to be correct* and therefore the derived identifications of the other cipher letters may be assumed to be correct. If the original starting point for assignment of plain-text values is not correct, or if the direction of "reading" the successive crests and troughs of the distribution is not correct, then the frequencies of the other 25 cipher letters will not correspond to or even approximate the normal theoretical frequencies of their hypothetical plain-text equivalents on the basis of the initial identification. A new initial point, that is, a different cipher equivalent must then be selected to represent E_p; or else the direction of "reading" the crests and troughs must be reversed. This procedure, that is, the attempt to make the actual frequency relations exhibited by uniliteral frequency distribution for a given cryptogram conform to the theoretical frequency relations of the normal frequency distribution in an effort to solve the cryptogram, is referred to technically as "fitting the actual uniliteral frequency distribution for a cryptogram to the theoretical uniliteral frequency distribution for normal plain text", or, more briefly, as *"fitting the frequency distribution for the cryptogram to the normal frequency distribution"*, or, still more briefly, *"fitting the distribution to the normal."* In statistical work the expression commonly employed in connection with this process of fitting an actual distribution to a theoretical one is "testing the goodness of fit." The goodness of fit may be stated in various ways, mathematical in character.

c. In fitting the actual distribution to the normal, it is necessary to regard the cipher component (that is, the letters A . . . Z marking the successive crests and troughs of the distribution) as partaking of the nature of a wheel or sequence closing in upon itself, so that no matter with what crest or trough one starts, the spatial and frequency relations of the crests and troughs are constant. This manner of regarding the cipher component as being cyclic in nature is valid *because it is obvious that the relative positions and frequencies of the crests and troughs of any uniliteral-frequency distribution must remain the same regardless of what letter is employed as the initial point of the distribution.* Fig. 5 gives a clear picture of what is meant in this connection, as applied to the normal frequency distribution.

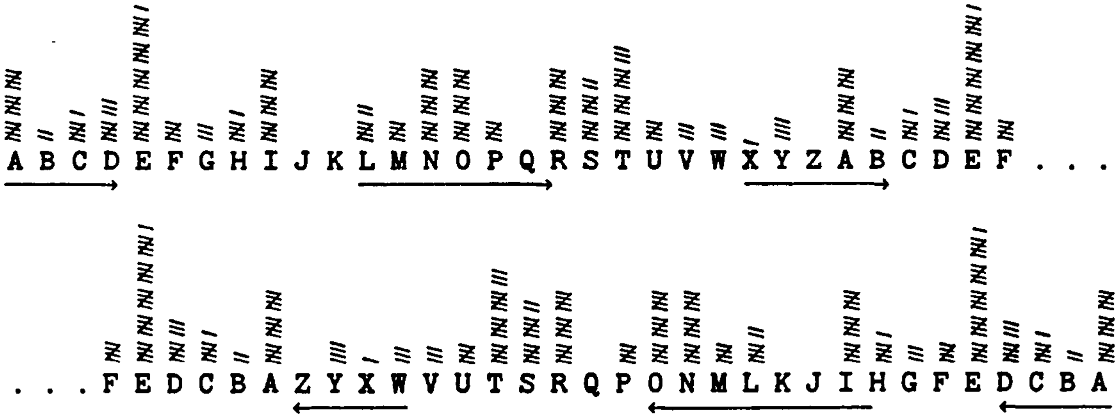


FIGURE 5.

d. In the third sentence of subparagraph *b*, the phrase "assumed to be correct" was advisedly employed in describing the results of the attempt to fit the distribution to the normal, because the final test of the goodness of fit in this connection (that is, of the correctness of the assignment of values to the crests and troughs of the distribution) is whether the *consistent* substitution of the plain-text values of the cipher characters in the cryptogram will yield intelligible plain text. If this is not the case, then no matter how close the approximation between actual and theoretical frequencies is, no matter how well the actual frequency distribution fits the normal, the only possible inferences are that (1) either the closeness of the fit is a pure coincidence in this case, and that another equally good fit may be obtained from the same data, or else (2) the cryptogram involves something more than simple monoalphabetic substitution by means of a single standard cipher alphabet. For example, suppose a transposition has been applied in addition to the substitution. Then, although an excellent correspondence between the uniliteral frequency distribution and the normal frequency distribution has been obtained, the substitution of the cipher letters by their assumed equivalents will still not yield plain text. However, aside from such cases of double encipherment, instances in which the uniliteral frequency distribution may be easily fitted to the normal frequency distribution and in which at the same time an attempted simple substitution fails to yield intelligible text are rare. It may be said that, in practical operations whenever the uniliteral frequency distribution can be made to fit the normal frequency distribution, substitution of values will result in solution; and, as a corollary, whenever the uniliteral frequency distribution cannot be made to fit the normal frequency distribution, the cryptogram does not represent a case of simple, monoalphabetic substitution by means of a standard alphabet.

18. Theoretical example of solution.—*a.* The foregoing principles will become clearer by noting the cryptographing and solution of a theoretical example. The following message is to be cryptographed.

HOSTILE FORCE ESTIMATED AT ONE REGIMENT INFANTRY AND TWO PLATOONS CAVALRY MOVING SOUTH ON QUINNIMONT PIKE STOP HEAD OF COLUMN NEARING ROAD JUNCTION SEVEN THREE SEVEN COMMA EAST OF GREENACRE SCHOOL FIRED UPON BY OUR PATROLS STOP HAVE DESTROYED BRIDGE OVER INDIAN CREEK.

b. First, solely for purposes of demonstrating certain principles, the uniliteral frequency distribution for this message is presented in Figure 6.

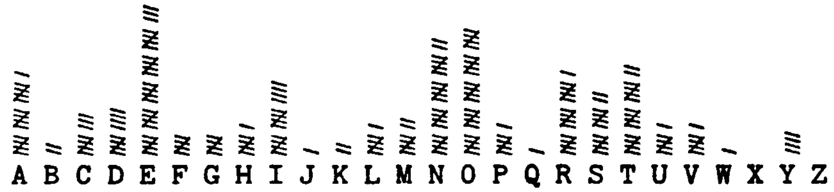


FIGURE 6.

c. Now let the foregoing message be cryptographed monoalphabetically by the following cipher alphabet, yielding the cryptogram and the frequency distribution shown below.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F

Plain	HOSTI	LEFOR	CEEST	IMATE	DATON	EREGI	MENTI	NFANT	RYAND
Cipher	NUYZO	RKLUX	IKKYZ	OSGZK	JGZUT	KXKMO	SKTZO	TLGTZ	XEGTJ
Plain	TWOPL	ATOON	SCAVA	LRYMO	VINGS	OUTHO	NQUIN	NIMON	TPIKE
Cipher	ZCUVR	GZUUT	YIGBG	RXESU	BOTMY	UAZNU	TWAOT	TOSUT	ZVOQK
Plain	STOPH	EADOF	COLUM	NNEAR	INGRO	ADJUN	CTION	SEVEN	THREE
Cipher	YZUVN	KGJUL	IURAS	TTKGX	OTMXU	GJPAT	IZOUT	YKBKT	ZNXKK
Plain	SEVEN	COMMA	EASTO	FGREE	NACRE	SCHOO	LFIRE	DUPON	BYOUR
Cipher	YKBKT	IUSSG	KGYZU	LMXKK	TGIXK	YINUU	RLOXK	JAVUT	HEUAX
Plain	PATRO	LSSTO	PHAVE	DESTR	OYEDB	RIDGE	OVERI	NDIAN	CREEK
Cipher	VGZXU	RYYZU	VNGBK	JKYZX	UEKJH	XOJMK	UBKXO	TJOGT	IXKKQ

CRYPTOGRAM

```
NUYZO  RKLUX  IKKYZ  OSGZK  JGZUT  KXKMO
SKTZO  TLGTZ  XEGTJ  ZCUVR  GZUUT  YIGBG
RXESU  BOTMY  UAZNU  TWAOT  TOSUT  ZVOQK
YZUVN  KGJUL  IURAS  TTKGX  OTMXU  GJPAT
IZOUT  YKBKT  ZNXKK  YKBKT  IUSSG  KGYZU
LMXKK  TGIXK  YINUU  RLOXK  JAVUT  HEUAX
VGZXU  RYYZU  VNGBK  JKYZX  UEKJH  XOJMK
UBKXO  TJOGT  IXKKQ
```

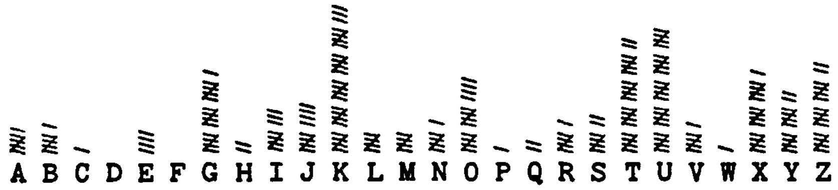


FIGURE 7

d. Let the student now compare Figs. 6 and 7, which have been superimposed in Fig. 8 for convenience in examination. Crests and troughs are present in both distributions; moreover their relative positions and frequencies have not been changed in the slightest particular. Only the absolute position of the sequence as a whole has been displaced six intervals to the right in Fig. 7, as compared with the absolute position of the sequence in Fig. 6.

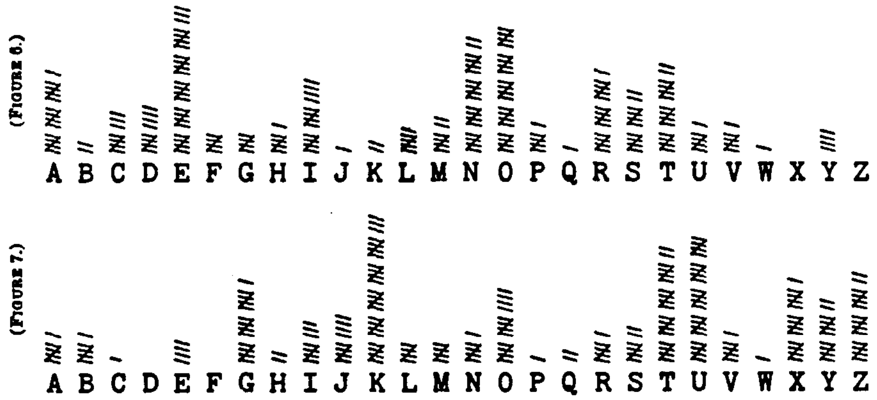


FIGURE 8.

e. If the two distributions are compared in detail the student will clearly understand how easy the solution of the cryptogram would be to one who knew nothing about how it was prepared. For example, the frequency of the highest crest, representing E_p in Fig. 6 is 28; at an interval of four letters before E_p there is another crest representing A_p with frequency 16. Between A and E there is a trough, representing the low-frequency letters B, C, D. On the other side of E, at an interval of four letters, comes another crest, representing I with frequency 14. Between E and I there is another trough, representing the low-frequency letters F, G, H. Compare these crests and troughs with their homologous crests and troughs in Fig. 7. In the latter, the letter K marks the highest crest in the distribution with a frequency of 28; four letters before K there is another crest, frequency 16, and four letters on the other side of K there is another crest, frequency

14. Troughs corresponding to B, C, D and F, G, H are seen at H, I, J and L, M, N in Fig. 7. In fact, the two distributions may be made to coincide exactly, by shifting the frequency distribution for the cryptogram six intervals to the left with respect to the distribution for the equivalent plain-text message, as shown herewith.

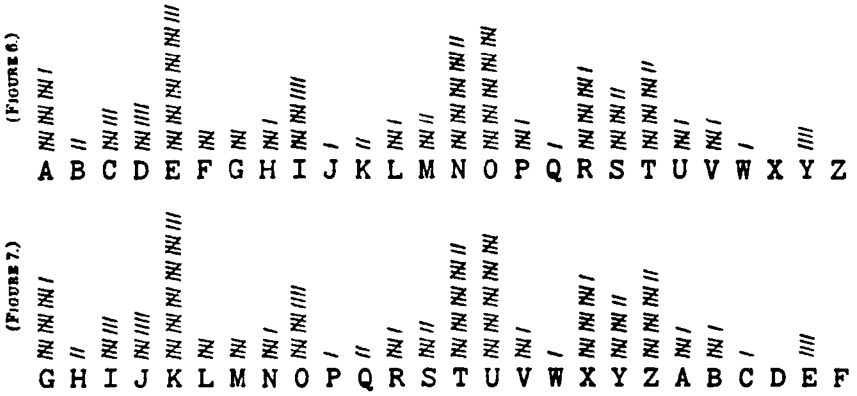


FIGURE 9.

f. Let us suppose now that nothing is known about the cryptographing process, and that only the cryptogram and its uniliteral frequency distribution is at hand. It is clear that simply bearing in mind the spatial relations of the crests and troughs in a normal frequency distribution would enable the cryptanalyst to fit the distribution to the normal in this case. He would naturally first assume that $G_c=A_p$, from which it would follow that if a direct standard alphabet is involved, $H_c=B_p$, $I_c=C_p$, and so on, yielding the following (tentative) deciphering alphabet:

Cipher............	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Plain.............	U V W X Y Z A B C D E F G H I J K L M N O P Q R S T

g. Now comes the final test: If these assumed values are substituted in the cipher text, the plain text immediately appears. Thus:

N U Y Z O	R K L U X	I K K Y Z	O S G Z K	J G Z U T	etc.
H O S T I	L E F O R	C E E S T	I M A T E	D A T O N	etc.

h. It should be clear, therefore, that the selection of G_c to represent A_p in the cryptographing process has absolutely no effect upon the relative spatial and frequency relations of the crests and troughs of the frequency distribution for the cryptogram. If Q_c had been selected to represent A_p, these relations would still remain the same, the whole series of crests and troughs being merely displaced further to the right of the positions they occupy when $G_c=A_p$.

19. Practical example of solution by the frequency method.—*a. The case of direct standard alphabet ciphers.*—(1) The following cryptogram is to be solved by applying the foregoing principles:

```
I B M Q O   P B I U O   M B B G A   J C Z O F   M U U Q B   A J C Z O
Z W I L N   Q T T M L   E Q B P U   I Z K P Q   V O Q V N   I V B Z G
```

(2) From the presence of repetitions and so many low-frequency letters such as B, Q, and Z it is at once suspected that this is a substitution cipher. But to illustrate the steps that must be taken in difficult cases in order to be certain in this respect, a uniliteral frequency distribution

is constructed, and then reference is made to charts 1 to 4 to note whether the actual numbers of vowels, high, medium, and low-frequency consonants fall inside or outside the areas delimited by the respective curves.

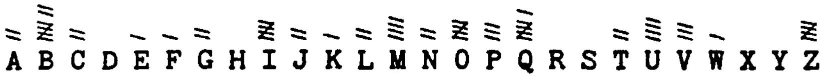


FIGURE 10*a*.

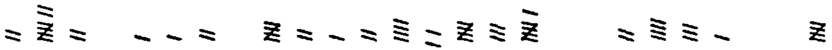
Letters	Frequency	Position with respect to areas delimited by curves
Vowels (A E I O U Y)	17	Outside, chart 1.
High-frequency Consonants (D N R S T)	4	Outside, chart 2.
Medium-frequency Consonants (B C F G H L M P V W)	25	Outside, chart 3.
Low-frequency Consonants (J K Q X Z)	14	Outside, chart 4.
Total	60	

(3) All four points falling quite outside the areas delimited by the curves applicable to these four classes of letters, the cryptogram is clearly a substitution cipher.

(4) The appearance of the frequency distribution, with marked crests and troughs, indicates that the cryptogram is probably monoalphabetic. Reference is now made to Chart 5. The message has 60 letters and 6 blanks. The point of intersection on the chart is closer to curve *P* than it is to curve *R*; therefore, this is additional evidence that the message is probably monoalphabetic.

(5) The next step is to determine whether a standard or a mixed cipher alphabet is involved. This is done by studying the positions and the sequence of crests and troughs in the frequency distribution, and trying to fit the distribution to the normal.

(6) The first assumption to be made is that a direct standard is involved. The highest crest in the distribution is marked by B_c. Let it be assumed that $B_c=E_p$. Then C_c, D_c, E_c, . . . $=F_p$, G_p, H_p, . . ., respectively; thus:



Cipher........... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Plain.............. D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

FIGURE 10*b*.

At first glance the approximation to the expected frequencies seems fair, especially in the region F G H I J K_p and R S T_p. But there are too many occurrences of L_p, P_p, X_p and C_p and too few occurrences of A_p, I_p, N_p, O_p. Moreover, if a substitution is attempted on this basis, the following is obtained for the first two cipher groups:

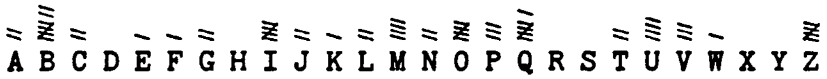
Cipher.......................... I B M Q O P B I U O
"Plain text"................... L E P T R S E L X R

This is certainly not plain text and it seems clear that B_c is not E_p. A different assumption will have to be made.

(7) Suppose $Q_c=E_p$. Going through the same steps as before, again no satisfactory results are obtained. Further trials[1] are made along the same lines, until the assumption $M_c=E_p$ is tested.

---

[1] It is unnecessary, of course, to write out the alphabets as shown in Figs. 10*b* and *c* when testing assumptions. This is usually all done mentally.



Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R

FIGURE 10c.

(8) The fit in this case is quite good; possibly there are too many occurrences of G_p and M_p and two few of E_p, O_p and S_p. But the final test remains: trial of the substitution alphabet on the cryptogram itself. This is immediately done and the results are as follows:

Cryptogram	IBMQO	PBIUO	NBBGA	JCZOF	MUUQB	AJCZO
Plain text	ATEIG	HTAMG	ETTYS	BURGX	EMMIT	SBURG
Cryptogram	ZWILN	QTTML	EQBPU	IZKPQ	VOQVN	IVBZG
Plain text	ROADF	ILLED	WITHM	ARCHI	NGINF	ANTRY

AT EIGHT AM GETTYSBURG—EMMITSBURG ROAD FILLED WITH MARCHING INFANTRY.

(9) It is always advisable to note the specific key. In this case the correspondence between any plain-text letter and its cipher equivalent will indicate the key. Although other conventions are possible, and equally valid, it is usual, however, to indicate the key by noting the cipher equivalent of A_p. In this case $A_p = I_c$.

b. The case of reversed standard alphabet ciphers.—

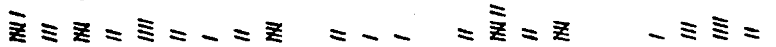
(1) Let the following cryptogram and its frequency distribution be studied.

I P E A C   B P I W C   E P P K Q   H O R C L   E W W A P   Q H O R C
R U I F D   A X X E F   M A P B W   I R G B A   V C A V D   I V P R K

(2) The preliminary steps illustrated above, under subpar. *a* (1) to (4) inclusive, in connection with the test for class and monoalphabeticity, will here be omitted, since they are exactly the same in nature. The result is that the cryptogram is obviously a substitution cipher and is monoalphabetic.

(3) Assuming that it is not known whether a direct or a reversed standard alphabet is involved, attempts are at once made to fit the frequency distribution to the normal direct sequence. If the student will try them he will soon find out that these are unsuccessful. All this takes but a few minutes.

(4) The next logical assumption is now made, viz, that the cipher alphabet is a reversed standard alphabet. When on this basis E_c is assumed to be E_p, the distribution can readily be fitted to the normal, practically every crest and trough in the actual distribution corresponding to a crest or trough in the expected distribution.



Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	I	H	G	F	E	D	C	B	A	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K	J

FIGURE 10d.

(5) When the substitution is made in the cryptogram, the following is obtained.

Cryptogram	IPEAC	BPIWC	EPPKQ	. . .
Plain text	ATEIG	HTAMG	ETTYS	. . .

(6) The plain-text message is identical with that under paragraph *a*. The specific key in this case is also $A_p = I_c$. If the student will compare the frequency distributions in the two cases,

he will note that the relative positions and extensions of the crests and troughs are identical; they merely progress in opposite directions.

20. **Solution by completing the plain-component sequence.**—*a. The case of direct standard alphabet ciphers.*—(1) The foregoing method of analysis, involving as it does the construction of a uniliteral frequency distribution, was termed a *solution by the frequency method* because it involves the construction of a frequency distribution and its study. There is, however, another method which is much more rapid, almost wholly mechanical, and which, moreover, does not necessitate the construction or study of any frequency distribution whatever. An understanding of the method follows from a consideration of the method of encipherment of a message by the use of a single, direct standard cipher alphabet.

(2) Note the following encipherment:

Message.......... REPEL INVADING CAVALRY

ENCIPHERING ALPHABET

Plain..............	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher...........	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F

ENCIPHERMENT

Plain text.......	R E P E L	I N V A D I N G	C A V A L R Y
Cryptogram....	X K V K R	O T B G J O T M	I G B G R X E

CRYPTOGRAM

X K V K R O T B G J O T M I G B G R X E

(3) The enciphering alphabet shown above represents a case wherein the sequence of letters of both components of the cipher alphabet is the normal sequence, with the sequence forming the cipher component merely shifted six intervals in retard (or 20 intervals in advance) of the position it occupies in the normal alphabet. If, therefore, two strips of paper bearing the letters of the normal sequence, equally spaced, are regarded as the two components of the cipher alphabet and are juxtaposed at all of the 25 possible points of coincidence, it is obvious that one of these 25 juxtapositions *must* correspond to the actual juxtaposition shown in the enciphering alphabet directly above.[2] It is equally obvious that if a record were kept of the results obtained by applying the values given at each juxtaposition to the letters of the cryptogram, one of these results would yield the plain text of the cryptogram.

(4) Let the work be systematized and the results set down in an orderly manner for examination. It is obviously unnecessary to juxtapose the two components so that $A_c=A_p$, for on the assumption of a direct standard alphabet, juxtaposing two direct normal components at their normal point of coincidence merely yields plain text. The next possible juxtaposition, therefore, is $A_c=B_p$. Let the juxtaposition of the two sliding strips therefore be $A_c=B_p$, as shown here:

```
Plain................ ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher..............   ABCDEFGHIJKLMNOPQRSTUVWXYZ
```

The values given by this juxtaposition are substituted for the first 20 letters of the cryptogram and the following results are obtained.

Cryptogram..................	X K V K R	O T B G J	O T M I G	B G R X E
1st Test—"Plain text"....	Y L W L S	P U C H K	P U N J H	C H S Y F

---

[2] One of the strips should bear the sequence repeated. This permits juxtaposing the two sequences at all 26 possible points of coincidence so as to have a complete cipher alphabet showing at all times.

This certainly is not intelligible text; obviously, the two components were not in the position indicated in this first test. The cipher component is therefore slid one interval to the right, making $A_c = C_p$, and a second test is made. Thus

Plain.................. ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher............... ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cryptogram..................... X K V K R O T B G J O T M I G B G R X E
2d Test—"Plain text"..... Z M X M T Q V D I L Q V O K I D I T Z G

Neither does the second test result in disclosing any plain text. But, if the results of the two tests are studied a phenomenon that at first seems quite puzzling comes to light. Thus, suppose the results of the two tests are superimposed in this fashion.

Cryptogram.....................	X K V K R	O T B G J	O T M I G	B G R X E
1st Test—"Plain text"....	Y L W L S	P U C H K	P U N J H	C H S Y F
2nd Test—"Plain text"...	Z M X M T	Q V D I L	Q V O K I	D I T Z G

(5) Note what has happened. The net result of the two experiments was merely to continue the normal sequence begun by the cipher letters at the heads of the several *columns*. It is obvious that if the normal sequence is completed in each column *the results will be exactly the same as though the whole set of 25 possible tests had actually been performed*. Let the columns therefore be completed, as shown in Fig. 11.

X	K	V	K	R	O	T	B	G	J	O	T	M	I	G	B	G	R	X	E
Y	L	W	L	S	P	U	C	H	K	P	U	N	J	H	C	H	S	Y	F
Z	M	X	M	T	Q	V	D	I	L	Q	V	O	K	I	D	I	T	Z	G
A	N	Y	N	U	R	W	E	J	M	R	W	P	L	J	E	J	U	A	H
B	O	Z	O	V	S	X	F	K	N	S	X	Q	M	K	F	K	V	B	I
C	P	A	P	W	T	Y	G	L	O	T	Y	R	N	L	G	L	W	C	J
D	Q	B	Q	X	U	Z	H	M	P	U	Z	S	O	M	H	M	X	D	K
E	R	C	R	Y	V	A	I	N	Q	V	A	T	P	N	I	N	Y	E	L
F	S	D	S	Z	W	B	J	O	R	W	B	U	Q	O	J	O	Z	F	M
G	T	E	T	A	X	C	K	P	S	X	C	V	R	P	K	P	A	G	N
H	U	F	U	B	Y	D	L	Q	T	Y	D	W	S	Q	L	Q	B	H	O
I	V	G	V	C	Z	E	M	R	U	Z	E	X	T	R	M	R	C	I	P
J	W	H	W	D	A	F	N	S	V	A	F	Y	U	S	N	S	D	J	Q
K	X	I	X	E	B	G	O	T	W	B	G	Z	V	T	O	T	E	K	R
L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T
N	A	L	A	H	E	J	R	W	Z	E	J	C	Y	W	R	W	H	N	U
O	B	M	B	I	F	K	S	X	A	F	K	D	Z	X	S	X	I	O	V
P	C	N	C	J	G	L	T	Y	B	G	L	E	A	Y	T	Y	J	P	W
Q	D	O	D	K	H	M	U	Z	C	H	M	F	B	Z	U	Z	K	Q	X
*R	E	P	E	L	I	N	V	A	D	I	N	G	C	A	V	A	L	R	Y
S	F	Q	F	M	J	O	W	B	E	J	O	H	D	B	W	B	M	S	Z
T	G	R	G	N	K	P	X	C	F	K	P	I	E	C	X	C	N	T	A
U	H	S	H	O	L	Q	Y	D	G	L	Q	J	F	D	Y	D	O	U	B
V	I	T	I	P	M	R	Z	E	H	M	R	K	G	E	Z	E	P	V	C
W	J	U	J	Q	N	S	A	F	I	N	S	L	H	F	A	F	Q	W	D

FIGURE 11.

An examination of the successive horizontal lines of the diagram discloses *one and only one* line of plain text, that marked by the asterisk and reading **REPELINVADINGCAVALRY.**

(6) Since each column in Fig. 11 is nothing but a normal sequence, it is obvious that instead of laboriously writing down these columns of letters every time a cryptogram is to be examined, it would be more convenient to prepare a set of strips each bearing the normal sequence doubled (to permit complete coincidence for an entire alphabet at any setting), and have them available for examining any future cryptograms. In using such a set of sliding strips in order to solve a cryptogram prepared by means of a single direct standard cipher alphabet, or to make a test to determine whether a cryptogram has been so prepared, it is only necessary to "set up" the letters of the cryptogram on the strips, that is, align them in a single row across the strips (by sliding the individual strips up or down). The successive horizontal lines, called *generatrices* (singular, *generatrix*), are then examined in a search for intelligible text. If the cryptogram really belongs to this simple type of cipher, one of the generatrices will exhibit intelligible text all the way across; this text will practically invariably be the plain text of the message. This method of analysis may be termed *a solution by completing the plain-component sequence*. Sometimes it is referred to as "running down" the sequence. The principle upon which the method is based constitutes one of the cryptanalyst's most valuable tools.[3]

b. The case of reversed standard alphabets.—(1) The method described under subpar. *a* may also be applied, in slightly modified form, in the case of a cryptogram enciphered by a single reversed standard alphabet. The basic principles are identical in the two cases.

(2) To show this it is necessary to experiment with two sliding components as before, except that in this case one of the components must be a reversed normal sequence, the other, a direct normal sequence.

(3) Let the two components be juxtaposed A to A, as shown below, and then let the resultant values be substituted for the letters of the cryptogram. Thus:

CRYPTOGRAM

```
                   P C R C V   Y T L G D   Y T A E G   L G V P I
Plain............. ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher............  ZYXWVUTSRQPONMLKJIHGFEDCBA

Cryptogram.................. P C R C V   Y T L G D   Y T A E G   L G V P I
1st Test—"Plain text"....... L Y J Y F   C H P U X   C H A W U   P U F L S
```

(4) This does not yield intelligible text, and therefore the reversed component is slid one space forward and a second test is made. Thus:

```
Plain............. ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher............   ZYXWVUTSRQPONMLKJIHGFEDCBA

Cryptogram.................. P C R C V   Y T L G D   Y T A E G   L G V P I
2d Test—"Plain text"........ M Z K Z G   D I Q V Y   D I B X V   Q V G M T
```

(5) Neither does the second test yield intelligible text. But let the results of the two tests be superimposed. Thus:

```
Cryptogram.................. P C R C V   Y T L G D   Y T A E G   L G V P I
1st Test—"Plain text"....... L Y J Y F   C H P U X   C H A W U   P U F L S
2d Test—"Plain text"........ M Z K Z G   D I Q V Y   D I B X V   Q V G M T
```

---

[3] It is recommended that the student prepare a set of 25 strips ¼ by ½ by 15 inches, made of well-seasoned wood, and glue alphabet strips to the wood. The alphabet on each strip should be a double or repeated alphabet with all letters equally spaced.

(6) It is seen that the letters of the "plain text" given by the *second* trial are merely the continuants of the normal sequences initiated by the letters of the "plain text" given by the first trial. If these sequences are "run down"—that is, completed within the columns—the results must obviously be the same as though successive tests exactly similar to the first two were applied to the cryptogram, using one reversed normal and one direct normal component. If the cryptogram has really been prepared by means of a single reversed standard alphabet, one of the generatrices of the diagram that results from completing the sequences *must* yield intelligible text.

(7) Let the diagram be made, or better yet, if the student has already at hand the set of sliding strips referred to in the footnote to page 36, let him "set up" the letters given by the *first* trial. Fig. 12 shows the diagram and indicates the plain-text generatrix.

P	C	R	C	V	Y	T	L	G	D	Y	T	A	E	G	L	G	V	P	I
L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T
N	A	L	A	H	E	J	R	W	Z	E	J	C	Y	W	R	W	H	N	U
O	B	M	B	I	F	K	S	X	A	F	K	D	Z	X	S	X	I	O	V
P	C	N	C	J	G	L	T	Y	B	G	L	E	A	Y	T	Y	J	P	W
Q	D	O	D	K	H	M	U	Z	C	H	M	F	B	Z	U	Z	K	Q	X
*R	E	P	E	L	I	N	V	A	D	I	N	G	C	A	V	A	L	R	Y
S	F	Q	F	M	J	O	W	B	E	J	O	H	D	B	W	B	M	S	Z
T	G	R	G	N	K	P	X	C	F	K	P	I	E	C	X	C	N	T	A
U	H	S	H	O	L	Q	Y	D	G	L	Q	J	F	D	Y	D	O	U	B
V	I	T	I	P	M	R	Z	E	H	M	R	K	G	E	Z	E	P	V	C
W	J	U	J	Q	N	S	A	F	I	N	S	L	H	F	A	F	Q	W	D
X	K	V	K	R	O	T	B	G	J	O	T	M	I	G	B	G	R	X	E
Y	L	W	L	S	P	U	C	H	K	P	U	N	J	H	C	H	S	Y	F
Z	M	X	M	T	Q	V	D	I	L	Q	V	O	K	I	D	I	T	Z	G
A	N	Y	N	U	R	W	E	J	M	R	W	P	L	J	E	J	U	A	H
B	O	Z	O	V	S	X	F	K	N	S	X	Q	M	K	F	K	V	B	I
C	P	A	P	W	T	Y	G	L	O	T	Y	R	N	L	G	L	W	C	J
D	Q	B	Q	X	U	Z	H	M	P	U	Z	S	O	M	H	M	X	D	K
E	R	C	R	Y	V	A	I	N	Q	V	A	T	P	N	I	N	Y	E	L
F	S	D	S	Z	W	B	J	O	R	W	B	U	Q	O	J	O	Z	F	M
G	T	E	T	A	X	C	K	P	S	X	C	V	R	P	K	P	A	G	N
H	U	F	U	B	Y	D	L	Q	T	Y	D	W	S	Q	L	Q	B	H	O
I	V	G	V	C	Z	E	M	R	U	Z	E	X	T	R	M	R	C	I	P
J	W	H	W	D	A	F	N	S	V	A	F	Y	U	S	N	S	D	J	Q
K	X	I	X	E	B	G	O	T	W	B	G	Z	V	T	O	T	E	K	R

FIGURE 12.

(8) The only difference in procedure between this case and the preceding one (where the cipher alphabet was a direct standard alphabet) is that the letters of the cipher text are first "deciphered" by means of *any* reversed standard alphabet and then the columns are "run down", according to the normal A B C . . . Z sequence. For reasons which will become apparent very soon, the first step in this method is technically termed *converting the cipher letters into their plain-component equivalents;* the second step is the same as before, *viz, completing the plain-component sequence.*

21. Special remarks on the method of solution by completing the plain-component sequence.—*a*. The terms employed to designate the steps in the solution set forth in Par. 20*b*, viz, "converting the cipher letters into their plain-component equivalents" and "completing the plain-component sequence", accurately describe the process. Their meaning will become more clear as the student progresses with the work. It may be said that whenever the plain component of a cipher alphabet is a *known* sequence, no matter how it is composed, the difficulty and time required to solve any cryptogram involving the use of that plain component is practically cut in half. *In some cases this knowledge facilitates, and in other cases is the only thing that makes possible the solution of a very short cryptogram that might otherwise defy solution.* Later on an example will be given to illustrate what is meant in this regard.

b. The student should take note, however, of two qualifying expressions that were employed in a preceding paragraph to describe the results of the application of the method. It was stated that "one of the generatrices will exhibit intelligible text *all the way across;* this text will *practically invariably* be the plain text." Will there ever be a case in which more than one generatrix will yield intelligible text throughout its extent? That obviously depends almost entirely on the number of letters that are aligned to form a generatrix. If a generatrix contains but a very few letters, only five, for example, it may happen as a result of pure chance that there will be two or more generatrices showing what might be "intelligible text." Note in Fig. 11, for example, that there are several cases in which 3-letter and 4-letter English words (ANY, VAIN, GOT, TIP, etc.) appear on generatrices that are not correct, these words being formed by pure chance. But there is not a single case, in this diagram, of a 5-letter or longer word appearing fortuitously, because obviously the longer the word the smaller the probability of its appearance purely by chance; and the probability that two generatrices of 15 letters each will both yield intelligible text along their entire length is exceedingly remote, so remote, in fact, that in practical cryptography such a case may be considered nonexistent.[4]

c. The student should observe that in reality there is no difference whatsoever in principle between the two methods presented in subpars. *a* and *b* of Par. 20. In the former the preliminary step of converting the cipher letters into their plain-component equivalents is apparently not present but in reality it is there. The reason for its apparent absence is that in that case the plain component of the cipher alphabet is identical in all respects with the cipher component, so that the cipher letters require no conversion, or, rather, they are identical with the equivalents that would result if they were converted on the basis $A_c=A_p$. In fact, if the solution process had been arbitrarily initiated by converting the cipher letters into their plain-component equivalents at the setting $A_c=O_p$, for example, and the cipher component slid one interval to the right thereafter, the results of the first and second tests of Par. 20*a* would be as follows:

Cryptogram	X	K	V	K	R	O	T	B	G	J	O	T	M	I	G	B	G	R	X	E
1st Test—"Plain text"	L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
2nd Test—"Plain text"	M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T

Thus, the foregoing diagram duplicates in every particular the diagram resulting from the first two tests under Par. 20*b*: a first line of cipher letters, a second line of letters derived from them but showing externally no relationship with the first line, and a third line derived immediately from the second line by continuing the direct normal sequence. This point is brought to attention only for the purpose of showing that a single, broad principle is the basis of the general method of solution by completing the plain-component sequence, and once the student has this firmly in mind he will have no difficulty whatsoever in realizing when the principle is applicable, what a powerful cryptanalytic tool it can be, and what results he may expect from its application in specific instances.

---

[4] A person with patience and an inclination toward the curiosities of the science might construct a text of 15 or more letters which would yield two "intelligible" texts on the plain-component completion diagram.

d. In the two foregoing examples of the application of the principle, the plain component was a normal sequence but it should be clear to the student, if he has grasped what has been said in the preceding subparagraph, that this component may be a mixed sequence which, if known (that is, if the sequence of letters comprising the sequence is known to the cryptanalyst), can be handled just as readily as can a plain component that is a normal sequence.

e. It is entirely immaterial at what points the plain and the cipher components are juxtaposed in the preliminary step of converting the cipher letters into their plain-component equivalents. For example, in the case of the reversed alphabet cipher solved in Par. 20*b*, the two components were arbitrarily juxtaposed to give the value A=A, but they might have been juxtaposed at any of the other 25 possible points of coincidence without in any way affecting the final result, *viz*, the production of one plain-text generatrix in the completion diagram.

22. Value of mechanical solution as a short cut.—*a.* It is obvious that the very first step the student should take in his attempts to solve an unknown cryptogram that is obviously a substitution cipher is to try the mechanical method of solution by completing the plain-component sequence, using the normal alphabet, first direct, then reversed. This takes only a very few minutes and is conclusive in its results. It saves the labor and trouble of constructing a frequency distribution in case the cipher is of this simple type. Later on it will be seen how certain variations of this simple type may also be solved by the application of this method. Thus, a very easy short cut to solution is afforded, which even the experienced cryptanalyst never overlooks in his first attack on an unknown cipher.

b. It is important now to note that *if neither of the two foregoing attempts is successful in bringing plain text to light and the cryptogram is quite obviously monoalphabetic in character, the cryptanalyst is warranted in assuming that the cryptogram involves a mixed cipher alphabet.*[5] The steps to be taken in attacking a cipher of the latter type will be discussed in the next section.

---

[5] There is but one other possibility, already referred to under Par. 17*d*, which involves the case where transposition and monoalphabetic substitution processes have been applied in successive steps. This is unusual, however, and will be discussed in its proper place.

SECTION VI

UNILITERAL SUBSTITUTION WITH MIXED CIPHER ALPHABETS

Paragraph
Basic reason for the low degree of cryptographic security afforded by monoalphabetic cryptograms involving standard cipher alphabets ---- 23
Preliminary steps in the analysis of a monoalphabetic, mixed-alphabet cryptogram ---- 24
Further data concerning normal plain text ---- 25
Preparation of the work sheet ---- 26
Triliteral-frequency distributions ---- 27
Classifying the cipher letters into vowels and consonants ---- 28
Further analysis of the letters representing vowels and consonants ---- 29
Substituting deduced values in the cryptogram ---- 30
Completing the solution ---- 31
General remarks on the foregoing solution ---- 32
The "probable-word" method; its value and applicability ---- 33
Solution of additional cryptograms produced by the same cipher component ---- 34

23. Basic reason for the low degree of cryptographic security afforded by monoalphabetic cryptograms involving standard cipher alphabets.—The student has seen that the solution of monoalphabetic cryptograms involving standard cipher alphabets is a very easy matter. Two methods of analysis were described, one involving the construction of a frequency distribution, the other not requiring this kind of tabulation, being almost mechanical in nature and correspondingly rapid. In the first of these two methods it was necessary to make a correct assumption as to the value of but one of the 26 letters of the cipher alphabet and the values of the remaining 25 letters at once become known; in the second method it was not necessary to assume a value for even a single cipher letter. The student should understand what constitutes the basis of this situation, *viz*, the fact that the two components of the cipher alphabet are composed of *known sequences*. What if one or both of these components are, for the cryptanalyst, *unknown sequences*? In other words, what difficulties will confront the cryptanalyst if the cipher component of the cipher alphabet is a mixed sequence? Will such an alphabet be solvable as a whole at one stroke, or will it be necessary to solve its values individually? Since the determination of the value of one cipher letter in this case gives no direct clues to the value of any other letter, it would seem that the solution of such a cipher should involve considerably more analysis and experiment than has the solution of either of the two types of ciphers so far examined occasioned. A typical example will be studied.

24. Preliminary steps in the analysis of a monoalphabetic, mixed alphabet cryptogram.—*a*. Note the following cryptogram:

```
SFDZF IOGHL PZFGZ DYSPF HBZDS GVHTF UPLVD FGYVJ VFVHT GADZZ AITYD
ZYFZJ ZTGPT VTZBD VFHTZ DFXSB GIDZY VTXOI YVTEF VMGZZ THLLV XZDFM
HTZAI TYDZY BDVFH TZDFK ZDZZJ SXISG ZYGAV FSLGZ DTHHT CDZRS VTYZD
OZFFH TZAIT YDZYG AVDGZ ZTKHI TYZYS DZGHU ZFZTG UPGDI XWGHX ASRUZ
DFUID EGHTV EAGXX
```

b. A casual inspection of the text discloses the presence of several long repetitions as well as of many letters of normally low frequency, such as F, G, V, X, and Z; on the other hand, letters of normally high frequency, such as the vowels, and the consonants N and R, are relatively scarce. The cryptogram is obviously a substitution cipher and the usual mechanical tests for determining whether it is possibly of the monoalphabetic, standard-alphabet type are applied. The results being negative, a uniliteral frequency distribution is immediately constructed and is as shown in Figure 13.

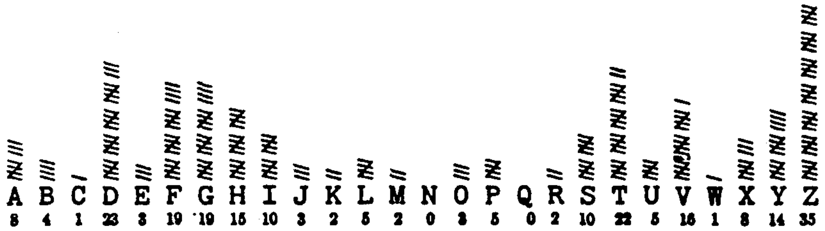


Figure 13

c. The fact that the frequency distribution shows very marked crests and troughs means that the cryptogram is undoubtedly monoalphabetic; the fact that it has already been tested (by the method of completing the plain-component sequence) and found not to be of the monoalphabetic, standard-alphabet type, indicates with a high degree of probability that it involves a mixed cipher alphabet. A few moments might be devoted to making a careful inspection of the distribution to insure that it cannot be made to fit the normal; the object of this would be to rule out the possibility that the text resulting from substitution by a standard cipher alphabet had not subsequently been transposed. But this inspection in this case is hardly necessary, in view of the presence of long repetitions in the message.[1] (See Par. 13*g*.)

d. One might, of course, attempt to solve the cryptogram by applying the simple principles of frequency. One might, in other words, assume that Z_c (the letter of greatest frequency) represents E_p, D_c (the letter of next greatest frequency) represents T_p, and so on. If the message were long enough this simple procedure might more or less quickly give the solution. But the message is relatively short and many difficulties would be encountered. Much time and effort would be expended unnecessarily, because it is hardly to be expected that in a message of only 235 letters the relative order of frequency of the various cipher letters should exactly coincide with, or even closely approximate the relative order of frequency of letters of normal plain text found in a count of 50,000 letters. *It is to be emphasized that the beginner must repress the natural tendency to place too much confidence in the generalized principles of frequency and to rely too much upon them.* It is far better to bring into effective use certain other data concerning normal plain text which thus far have not been brought to notice.

25. **Further data concerning normal plain text.**—*a*. Just as the individual letters constituting a large volume of plain text have more or less characteristic or fixed frequencies, so it is found that digraphs and trigraphs have characteristic frequencies, when a large volume of text is studied statistically. In Appendix 1, Table 6, are shown the relative frequencies of all digraphs appearing in the 260 telegrams referred to in Paragraph 9*e*. It will be noted that 428 of the 676 possible pairs of letters occur in these telegrams, but whereas many of them occur but once or twice, there are a few which occur hundreds of times.

b. In Appendix 1 will also be found several other kinds of tables and lists which will be useful to the student in his work, such as the relative order of frequency of the 50 digraphs of greatest frequency, the relative order of frequency of doubled letters, doubled vowels, doubled consonants, and so on. It is suggested that the student refer to this appendix now, to gain an idea of the data available for his future reference. Just how these data may be employed will become apparent very shortly.

---

[1] This possible step is mentioned here for the purpose of making it clear that the plain-component sequence completion method cannot solve a case in which transposition has followed or preceded monoalphabetic substitution with standard alphabets. Cases of this kind will be discussed in a later text. It is sufficient to indicate at this point that the frequency distribution for such a combined substitution-transposition cipher would present the characteristics of a standard alphabet cipher—and yet the method of completing the plain-component sequence would fail to bring out any plain text.

26. **Preparation of the work sheet.**—*a.* The details to be considered in this paragraph may at first appear to be superfluous but long experience has proved that systematization of the work, and preparation of the data in the most utilizable, condensed form is most advisable, even if this seems to take considerable time. In the first place if it merely serves to avoid interruptions and irritations occasioned by failure to have the data in an instantly available form, it will pay by saving mental wear and tear. In the second place, especially in the case of complicated cryptograms, painstaking care in these details, while it may not always bring about success, is often the factor that is of greatest assistance in ultimate solution. The detailed preparation of the data may be irksome to the student, and he may be tempted to avoid as much of it as possible, but, unfortunately, in the early stages of solving a cryptogram he does not know (nor, for that matter, does the expert always know) just which data are essential and which may be neglected. Even though not all of the data may turn out to have been necessary, as a general rule, time is saved in the end if all the usual data are prepared as a regular preliminary to the solution of most cryptograms.

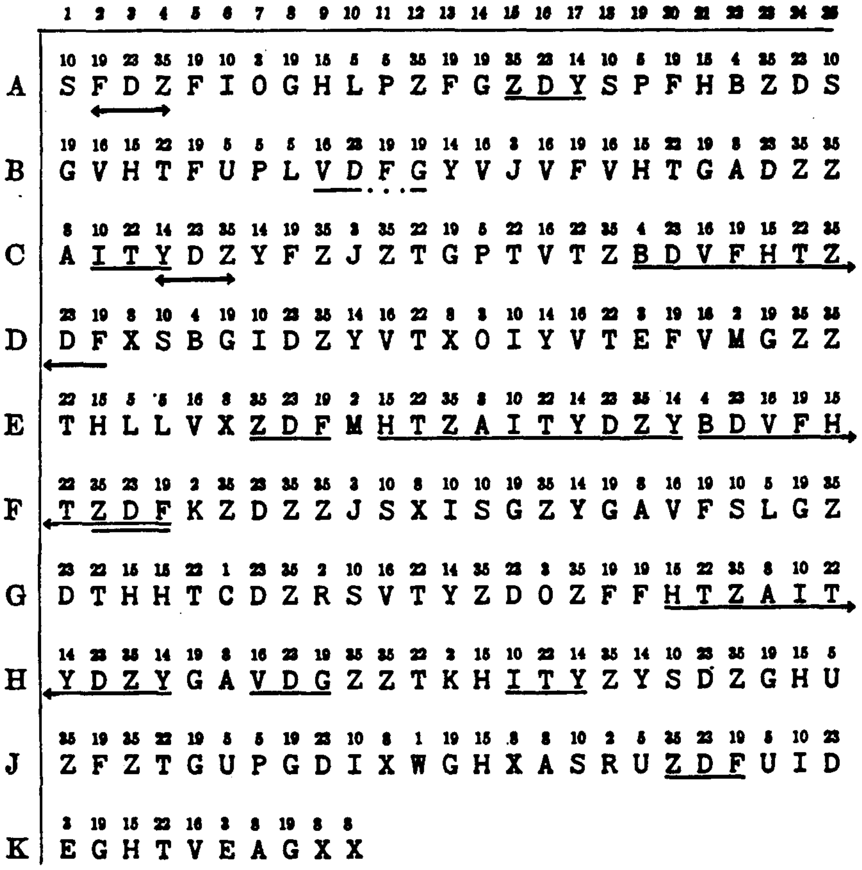
b. First, the cryptogram is recopied in the form of a *work sheet*. This sheet should be of a good quality of paper so as to withstand considerable erasure. If the cryptogram is to be copied by hand, cross-section paper of ¼-inch squares is extremely useful. The writing should be in ink, and plain, carefully made roman capital letters should be used in all cases. If the cryptogram is to be copied on a typewriter, the ribbon employed should be impregnated with an ink that will not smear or smudge under the hand.

c. The arrangement of the characters of the cryptogram on the work sheet is a matter of considerable importance. If the cryptogram as first obtained is in groups of regular length (usually five characters to a group) and if the uniliteral frequency distribution shows the cryptogram to be monoalphabetic, the characters should be copied without regard to this grouping. It is advisable to allow two spaces between letters, and to write a constant number of letters per line, approximately 25. At least two spaces, preferably three spaces, should be left between horizontal lines. Care should be taken to avoid crowding the letters in any case, for this is not only confusing to the eye but also mentally irritating when later it is found that not enough space has been left for making various sorts of marks or indications. If the cryptogram is originally in what appears to be word lengths (and this is the case, as a rule, only with the cryptograms of amateurs), naturally it should be copied on the work sheet in the original groupings. If further study of a cryptogram shows that some special grouping is required, it is often best to recopy it on a fresh work sheet rather than to attempt to indicate the new grouping on the old work sheet.

d. In order to be able to locate or refer to specific letters or groups of letters with speed, certainty, and without possibility of confusion, it is advisable to use coordinates applied to the lines and columns of the text as it appears on the work sheet. To minimize possibility of confusion, it is best to apply letters to the horizontal lines of the text, numbers to the vertical columns. In referring to a letter the horizontal line in which the letter is located is usually given first. Thus, referring to the work sheet shown below, coordinates A17 designate the letter Y, the 17th letter in the first line. The letter I is usually omitted from the series of line indicators so as to avoid confusion with the figure 1 If lines are limited to 25 letters each, then each set of 100 letters of the text is automatically blocked off by remembering that 4 lines constitute 100 letters.

e. Above each character of the cipher text may be some indication of the frequency of that character in the whole cryptogram. This indication may be the actual number of times the character occurs, or, if colored pencils are used, the cipher letters may be divided up into three categories or groups—high frequency, medium frequency, and low frequency. It is perhaps simpler, if clerical help is available, to indicate the actual frequencies. This saves constant reference to the frequency tables, which interrupts the train of thought, and saves considerable time in the end.

f. After the special frequency distribution, explained in Par. 27 below, has been constructed, repetitions of digraphs and trigraphs should be underscored. In so doing, the student should be particularly watchful of trigraphic repetitions which can be further extended into tetragraphs and polygraphs of greater length. Repetitions of more than ten characters should be set off by heavy vertical lines, as they indicate repeated phrases and are of considerable assistance in solution. If a repetition continues from one line to the next, put an arrow at the end of the underscore to signal this fact. Reversible digraphs should also be indicated by an underscore with an arrow pointing in both directions. Anything which strikes the eye as being peculiar, unusual, or significant as regards the distribution or recurrence of the characters should be noted. All these marks should, if convenient, be made with ink so as not to cause smudging. The work sheet will now appear as shown herewith (not all the repetitions are underscored):



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	10 S	19 F	23 D	35 Z	19 F	10 I	3 O	19 G	15 H	5 L	5 P	35 Z	19 F	19 G	35 Z	23 D	14 Y	10 S	5 P	19 F	15 H	4 B	35 Z	23 D	10 S
B	19 G	16 V	15 H	22 T	19 F	5 U	5 P	5 L	16 V	23 D	19 F	19 G	14 Y	16 V	3 J	16 V	19 F	16 V	15 H	22 T	19 G	8 A	23 D	35 Z	35 Z
C	8 A	10 I	22 T	14 Y	23 D	35 Z	14 Y	19 F	35 Z	3 J	35 Z	22 T	19 G	5 P	22 T	16 V	22 T	35 Z	4 B	23 D	16 V	19 F	15 H	22 T	35 Z
D	23 D	19 F	8 X	10 S	4 B	19 G	10 I	23 D	35 Z	14 Y	16 V	22 T	8 X	3 O	10 I	14 Y	16 V	22 T	3 E	19 F	16 V	2 M	19 G	35 Z	35 Z
E	22 T	15 H	5 L	5 L	16 V	8 X	35 Z	23 D	19 F	2 M	15 H	22 T	35 Z	8 A	10 I	22 T	14 Y	23 D	35 Z	14 Y	4 B	23 D	16 V	19 F	15 H
F	22 T	35 Z	23 D	19 F	2 K	35 Z	23 D	35 Z	35 Z	3 J	10 S	8 X	10 I	10 S	19 G	35 Z	14 Y	19 G	8 A	16 V	19 F	10 S	5 L	19 G	35 Z
G	23 D	22 T	15 H	15 H	22 T	1 C	23 D	35 Z	2 R	10 S	16 V	22 T	14 Y	35 Z	23 D	3 O	35 Z	19 F	19 F	15 H	22 T	35 Z	8 A	10 I	22 T
H	14 Y	23 D	35 Z	14 Y	19 G	8 A	16 V	23 D	19 G	35 Z	35 Z	22 T	2 K	15 H	10 I	22 T	14 Y	35 Z	14 Y	10 S	23 D	35 Z	19 G	15 H	5 U
J	35 Z	19 F	35 Z	22 T	19 G	5 U	5 P	19 G	23 D	10 I	8 X	1 W	19 G	15 H	8 X	8 A	10 S	2 R	5 U	35 Z	23 D	19 F	5 U	10 I	23 D
K	3 E	19 G	15 H	22 T	16 V	3 E	8 A	19 G	8 X	8 X															

27. **Triliteral-frequency distributions.**—*a*. In what has gone before, a type of frequency distribution known as a uniliteral frequency distribution was used. This, of course, shows only the number of times each individual letter occurs. In order to apply the normal digraphic and

trigraphic frequency data (given in Appendix 1) to the solution of a cryptogram of the type now being studied, it is obvious that the data with respect to digraphs and trigraphs occurring in the cryptogram should be compiled and should be compared with the data for normal plain text. In order to accomplish this in suitable manner, it is advisable to construct a slightly more complicated form of distribution termed a ***triliteral frequency distribution***.[3]

b. Given a cryptogram of 50 or more letters and the task of determining what trigraphs are present in the cryptogram, there are three ways in which the data may be arranged or assembled. One may require that the data show (1) each letter with its two succeeding letters; (2) each letter with its two preceding letters; (3) each letter with one preceding letter and one succeeding letter.

c. A distribution of the first of the three foregoing types may be designated as a "triliteral frequency distribution showing two suffixes"; the second type may be designated as a "triliteral frequency distribution showing two prefixes"; the third type may be designated as a "triliteral frequency distribution showing one prefix and one suffix." Quadriliteral and pentaliteral frequency distributions may occasionally be found useful.

d. Which of these three arrangements is to be employed at a specific time depends largely upon what the data are intended to show. For present purposes, in connection with the solution of a monoalphabetic substitution cipher employing a mixed alphabet, possibly the third arrangement, that showing one prefix and one suffix, is most satisfactory.

e. It is convenient to use ¼-inch cross-section paper for the construction of a triliteral frequency distribution in the form of a distribution showing crests and troughs, such as that in Figure 14. In that figure the prefix to each letter to be recorded is inserted in the left half of the cell directly above the cipher letter being recorded; the suffix to each letter is inserted in the right half of the cell directly above the letter being recorded; and in each case the prefix and the suffix to the letter being recorded occupy the same cell, the prefix being directly to the left of the suffix. The number in parentheses gives the total frequency for each letter.

---

[3] Heretofore such a distribution has been termed a "trigraphic frequency table." It is thought that the word "triliteral" is more suitable, to correspond with the designation "uniliteral" in the case of the distribution of the single letters. A trigraphic distribution of A B C D E F would consider only the trigraphs A B C and D E F, whereas a triliteral distribution would consider the groups A B C, B C D, C D E, and D E F. (See also Par. 11*d*.) The use of the word "distribution" to replace the word "table" has already been explained.

CONDENSED TABLE OF REPETITIONS

Digraphs			*Trigraphs*		*Longer Polygraphs*
DZ–9	TZ–5	VF–4	DZY–4	FHT–3	HTZAITYDZY–2
ZD–9	TY–5	VT–4	HTZ–4	TYD–3	BDVFHTZDF–2
HT–8	FH–4	ZF–4	ITY–4	YDZ–3	ZAITYDZY–3
ZY–6	GH–4	ZT–4	ZDF–4	ZAI–3	FHTZ–3
DF–5	IT–4	ZZ–4	AIT–3		
GZ–5					

																									UD
																									FT
																									UF
																									DG
																									YY
																									ZT
																									GZ
																									DY
																									TA
																									OF
																									YD
																									DR
			IE																						GD
			ZF																HV						GY
			GI																ZG						ZJ
			SZ																IY						DZ
			VG		DU	AX													ZK						KD
			YZ		ZZ	EH													IY						TD
			ZO		FH	WH													HZ						DY
			CZ		ZF	PD													VY		TE				TA
			ZT		VS	TU	GT												HC		AD				XD
			ZZ		DK	ZH	GX												DH		ST			ZS	ZT
			ZF		VH	DZ	GU												HZ		AF			TZ	GZ
			BV		DM	YA	KI												IY		DF			ZG	DY
			YZ		EV	LZ	FT												HZ		LX			TD	TD
			ZF		DX	YA	HT	UD										AR	ZH		FM			TZ	TB
			IZ		VH	SZ	TH	DX										YD	VE		YT			ZG	JT
EG			ZF		YZ	MZ	FT	HT										RV	VX		YT		X–	ZB	FJ
XS			BV		VV	BI	MT	AT										FL	HZ		DF		GX	TD	DY
GV			YZ		DG	TP	TL	XS										IG	VZ		TT		HA	IV	ZA
ZI			AZ		TU	TA	FT	AT			SG				UG			JX	PV	FI	FH		IW	ZV	DZ
GV	YD		VF		PH	FY	VT	OY			LV				GT			XB	ZG	RZ	JF		SI	ZF	BD
ZI	SG		ZS	VA	ZG	SV	VT	GD	ZS		HL			DZ	UL			DG	IY	GP	YJ		VZ	TD	GD
ZI	ZD		ZY	DG	ZI	FZ	FB	AT	ZZ	TH	PV	FH		XI	SF		SU	YP	HG	HZ	LD		TO	GV	PF
GD	HZ	TD	FZ	TF	SD	OH	GL	FO	VV	FZ	HP	VG		IG	LZ		ZS	–F	HF	FP	GH	XG	FS	DS	DF
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
(8)	(4)	(1)	(23)	(3)	(19)	(19)	(15)	(10)	(3)	(2)	(5)	(2)	(0)	(3)	(5)	(0)	(2)	(10)	(22)	(5)	(16)	(1)	(8)	(14)	(35)

FIGURE 14.

f. The triliteral frequency distribution is now to be examined with a view to ascertaining what digraphs and trigraphs occur two or more times in the cryptogram. Consider the pair of columns containing the prefixes and suffixes to D_c in the distribution, as shown in Fig. 14. This pair of columns shows that the following digraphs appear in the cryptogram:

Digraphs based on prefixes (arranged as one reads up the column)	*Digraphs based on suffixes (arranged as one reads up the column)*
FD, ZD, ZD, VD, AD, YD, BD,	DZ, DY, DS, DF, DZ, DZ, DV,
ZD, ID, ZD, YD, BD, ZD, ZD,	DF, DZ, DF, DZ, DV, DF, DZ,
ZD, CD, ZD, YD, VD, SD, GD,	DT, DZ, DO, DZ, DG, DZ, DI,
ZD, ID	DF, DE

The nature of the triliteral frequency distribution is such that in finding what digraphs are present in the cryptogram it is immaterial whether the prefixes or the suffixes to the cipher letters are studied, *so long as one is consistent in the study.* For example, in the foregoing list of digraphs based on the prefixes to D_c, the digraphs FD, ZD, ZD, VD, etc., are found; if now, the student will refer to the suffixes of F_c, Z_c, V_c, etc., he will find the very same digraphs indicated. This being the case, the question may be raised as to what value there is in listing both the prefixes and the suffixes to the cipher letters. The answer is that by so doing the trigraphs are indicated at the same time. For example, in the case of D_c, the following trigraphs are indicated:

FDZ, ZDY, ZDS, VDF, ADZ, YDZ, BDV, ZDF, IDZ, ZDF, YDZ, BDV, ZDF,
ZDZ, ZDT, CDZ, ZDO, YDZ, VDG, SDZ, GDI, ZDF, IDE.

g. The *repeated* digraphs and trigraphs can now be found quite readily. Thus, in the case of D_c, examining the list of digraphs based on suffixes, the following repetitions are noted:

DZ appears 9 times
DF appears 5 times
DV appears 2 times

Examining the trigraphs with D_c as central letter, the following repetitions are noted:

ZDF appears 4 times
YDZ appears 3 times
BDV appears 2 times

h. It is unnecessary, of course, to go through the detailed procedure set forth in the preceding subparagraphs in order to find all the repeated digraphs and trigraphs. The repeated trigraphs with D_c as central letter can be found merely from an inspection of the prefixes and suffixes opposite D_c in the distribution. It is necessary only to find those cases in which two or more prefixes are identical at the same time that the suffixes are identical. For example, the distribution shows at once that in four cases the prefix to D_c is Z_c at the same time that the suffix to this letter is F_c. Hence, the trigraph ZDF appears four times. The repeated trigraphs may all be found in this manner.

i. The most frequently repeated digraphs and trigraphs are then assembled in what is termed a *condensed table of repetitions,* so as to bring this information prominently before the eye. As a rule, digraphs which occur less than four or five times, and trigraphs which occur less than three or four times may be omitted from the condensed table as being relatively of no importance in the study of repetitions. In the condensed table the frequencies of the individual letters forming the most important digraphs, trigraphs, etc., should be indicated.

28. Classifying the cipher letters into vowels and consonants.—*a.* Before proceeding to a detailed analysis of the repeated digraphs and trigraphs, a very important step can be taken which will be of assistance not only in the analysis of the repetitions but also in the final solution of the cryptogram. This step concerns the classification of the high-frequency letters into two groups—vowels and consonants. For if the cryptanalyst can quickly ascertain the equivalents of the four vowels, A, E, I, and O, and of only the four consonants, N, R, S, and T, he will then have the values of approximately two-thirds of all the cipher letters that occur in the cryptogram; the values of the remaining letters can almost be filled in automatically.

b. The basis for the classification will be found to rest upon a comparatively simple phenomenon: the associational or combinatory behavior of vowels is, in general, quite different from that of consonants. If an examination be made of Table 7-B in Appendix 1, showing the relative order of frequency of the 18 digraphs composing 25 percent of English telegraphic text, it will be seen that the letter E enters into the composition of 9 of the 18 digraphs; that is, in exactly half of all the cases the letter E is one of the two letters forming the digraph. The digraphs containing E are as follows:

ED	EN	ER	ES	
	NE	RE	SE	TE	VE

The remaining nine digraphs are as follows:

AN	ND	OR	ST
IN	NT		TH
ON			TO

c. None of the 18 digraphs is a combination of vowels. Note now that of the 9 combinations with E, 7 are with the consonants N, R, S, and T, one is with D, one is with V, and *none is with any vowel.* In other words, E_p combines most readily with consonants but not with other vowels, or even with itself. Using the terms often employed in the chemical analogy, E shows a great "affinity" for the consonants N, R, S, T, but not for the vowels. Therefore, if the letters of highest frequency occurring in a given cryptogram are listed, together with the number of times each of them combines with the cipher equivalent of E_p, those which show considerable combining power or affinity for the cipher equivalent of E_p may be assumed to be the cipher equivalents of N, R, S, T_p; those which do not show any affinity for the cipher equivalent of E_p may be assumed to be the cipher equivalents of A, I, O, U_p. Applying these principles to the problem in hand, and examining the triliteral frequency distribution, it is quite certain that $Z_c=E_p$, not only because Z_c is the letter of highest frequency, but also because it combines with *several* other high-frequency letters, such as D_c, F_c, G_c, etc. The nine letters of next highest frequency are:

23	22	19	19	16	15	14	10	10
D	T	F	G	V	H	Y	S	I

Let the combinations these letters form with Z_c be indicated in the following manner:

Number of times Z_c occurs as prefix..	𝍸𝍷𝍷𝍷𝍷	𝍷𝍷𝍷𝍷	𝍷𝍷𝍷𝍷	𝍷			𝍸𝍷		
Cipher Letter..	D(23)	T(22)	F(19)	G(19)	V(16)	H(15)	Y(14)	S(10)	I(10)
Number of times Z_c occurs as suffix..	𝍸𝍷𝍷𝍷𝍷	𝍸	𝍷𝍷	𝍸			𝍷𝍷		

d. Consider D_c. It occurs 23 times in the message and 18 of those times it is combined with Z_c, 9 times in the form Z_cD_c $(=E\theta_p)$, and 9 times in the form D_cZ_c $(=\theta E_p)$. It is clear that D_c must be a consonant. In the same way, consider T_c, which shows 9 combinations with Z_c, 4 in the form Z_cT_c $(=E\theta_p)$ and 5 in the form T_cZ_c $(=\theta E_p)$. The letter T_c appears to represent a consonant, as do also the letters F_c, G_c, and Y_c. On the other hand, consider V_c, occurring in all 16 times but never in combination with Z_c; it appears to represent a vowel, as do also the letters H_c, S_c, and I_c. So far, then, the following classification would seem logical:

Vowels	*Consonants*
$Z_c(=E_p)$, V_c, H_c, S_c, I_c	D_c, T_c, F_c, G_c, Y_c

29. Further analysis of the letters representing vowels and consonants.—*a*. O_p is usually the vowel of second highest frequency. Is it possible to determine which of the letters V, H, S, I_c is the cipher equivalent of O_p? Let reference be made again to Table 6 in Appendix 1, where it is seen that the 10 most frequently occurring diphthongs are:

Diphthong	IO	OU	EA	EI	AI	IE	AU	EO	AY	UE
Frequency	41	37	35	27	17	13	13	12	12	11

If V, H, S, I_c are really the cipher equivalents of A, I, O, U_p (not respectively), perhaps it is possible to determine which is which *by examining the combinations they make among themselves and with* Z_c $(=E_p)$. Let the combinations of V, H, S, I, and Z that occur in the message be listed. There are only the following:

ZZ_c—4
VH —2
HH —1
HI—1
SV—1
IS—1

ZZ_c is of course EE_p. Note the doublet HH_c; if H_c is a vowel, then the chances are excellent that $H_c=O_p$ because the doublets AA_p, II_p, UU_p, are practically non-existent, whereas the double vowel combination OO_p is of next highest frequency to the double vowel combination EE_p. If $H_c=O_p$, then V_c must be I_p because the digraph VH_c occurring two times in the message could hardly be AO_p, or UO_p, whereas the diphthong IO_p is the one of high frequency in English. So far then, the tentative (because so far unverified) results of the analysis are as follows:

$$Z_c=E_p \quad H_c=O_p \quad V_c=I_p$$

This leaves only two letters, I_c and S_c (already classified as vowels) to be separated into A_p and U_p. Note the digraphs:

$$HI_c=O\theta_p$$
$$SV_c=\theta I_p$$
$$IS_c=\theta\theta_p$$

Only two alternatives are open:

(1) Either $I_c=A_p$ and $S_c=U_p$,
(2) Or $I_c=U_p$ and $S_c=A_p$.

If the first alternative is selected, then

$$HI_c=OA_p$$
$$SV_c=UI_p$$
$$IS_c=AU_p$$

If the second alternative is selected, then

$$HI_c=OU_p$$
$$SV_c=AI_p$$
$$IS_c=UA_p$$

The eye finds it difficult to choose between these alternatives; but suppose the frequency values of the plain-text diphthongs as given in Table 6 of Appendix 1 are added for each of these alternatives, giving the following:

$HI_c=OA_p$, frequency value= 7
$SV_c=UI_p$, frequency value= 5
$IS_c=AU_p$, frequency value=13
Total.............. 25

$HI_c=OU_p$, frequency value=37
$SV_c=AI_p$, frequency value=17
$IS_c=UA_p$, frequency value= 5
Total.............. 59

Mathematically, the second alternative is more than twice as probable as the first. Let it be assumed to be correct and the following (still tentative) values are now at hand:

$$Z_c=E_p \quad H_c=O_p \quad V_c=I_p \quad S_c=A_p \quad I_c=U_p$$

b. Attention is now directed to the letters classified as consonants*:* How far is it possible to ascertain their values? The letter D_c, from considerations of frequency alone, would seem to be T_p, but its frequency, 23, is not considerably greater than that for T_c. It is not much greater than that for F_c or G_c, with a frequency of 19 each. But perhaps it is possible to ascertain not the value of one letter alone but of two letters at one stroke. To do this one may make use of a tetragraph of considerable importance in English, *viz,* $TION_p$. For if the analysis pertaining to the vowels is correct, and if $VH_c=IO_p$, then an examination of the letters immediately before and after the digraph VH_c in the cipher text might disclose both T_p and N_p. Reference to the text gives the following:

$$GVHT_c \qquad FVHT_c$$
$$\theta IO\theta_p \qquad \theta IO\theta_p$$

The letter T_c follows VH_c in both cases and very probably indicates that $T_c=N_p$; but as to whether G_c or F_c equals T_p cannot be decided. However, two conclusions are clear: first, the letter D_c is neither T_p nor N_p, from which it follows that it must be either R_p or S_p; second, the letters G_c and F_c must be either T_p and S_p, respectively, or S_p and T_p, respectively, because the only tetragraphs usually found (in English) containing the diphthong IO_p as central letters are $SION_p$ and $TION_p$. This in turn means that as regards D_c, the latter cannot be *either* R_p or S_p; it *must* be R_p, a conclusion which is corroborated by the fact that ZD_c (=ER_p) and DZ_c (=RE_p) occur 9 times each. Thus far, then, the identifications, when inserted in an *enciphering* alphabet, are as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S				Z				V					T	H			D	G	F	I					
																			F	G						

30. Substituting deduced values in the cryptogram.—*a.* Thus far the analysis has been almost purely hypothetical, for as yet not a single one of the values deduced from the foregoing analysis has been tried out in the cryptogram. It is high time that this be done, because the final test of the validity of the hypotheses, assumptions, and identifications made in any cryptographic study is, after all, only this: do these hypotheses, assumptions, and identifications ultimately yield verifiable, intelligible plain-text when *consistently* applied to the cipher text?

b. At the present stage in the process, since there are at hand the assumed values of but 9 out of the 25 letters that appear, it is obvious that a continuous "reading" of the cryptogram can certainly not be expected from a mere insertion of the values of the 9 letters. However, the substitution of these values should do two things. First, it should immediately disclose the fragments, outlines, or "skeletons" of "good" words in the text; and second, it should disclose no places in the text where "impossible" sequences of letters are established. By the first is meant that the partially deciphered text should show the outlines or skeletons of words such as may be expected to be found in the communication; this will become quite clear in the next subparagraph. By the second is meant that sequences, such as "AOOEN" or "TNRSENO" or the like, obviously not possible or extremely unusual in normal English text, must not result from the substitution of the tentative identifications resulting from the analysis. The appearance of several such extremely unusual or impossible sequences at once signifies that one or more of the assumed values is incorrect.

c. Here are the results of substituting the nine values which have been deduced by the reasoning based on a classification of the high-frequency letters into vowels and consonants and the study of the members of the two groups:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	10 S A	19 F T S	23 D R	35 Z E	19 F T S	10 I	3 O	19 G S T	15 H O	5 L	5 P	35 Z E	19 F T S	19 G S T	35 Z E	23 D R	14 Y	10 S A	5 P	19 F T S	15 H O	4 B	35 Z E	23 D R	10 S A
B	19 G S T	16 V I	15 H O	22 T N	19 F T S	5 U	5 P	5 L	16 V I	23 D R	19 F T S	19 G S T	14 Y	16 V I	3 J	16 V I	19 F T S	16 V I	15 H O	22 T N	19 G S T	8 A	23 D R	35 Z E	35 Z E
C	8 A	10 I	22 T N	14 Y	23 D R	35 Z E	14 Y	19 F T S	35 Z E	3 J	35 Z E	22 T N	19 G S T	5 P	22 T N	16 V I	22 T N	35 Z E	4 B	23 D R	16 V I	19 F T S	15 H O	22 T N	35 Z E
D	23 D R	19 F T S	8 X	10 S A	4 B	19 G S T	10 I	23 D R	35 Z E	14 Y	16 V I	22 T N	8 X	3 O	10 I	14 Y	16 V I	22 T N	3 E	19 F T S	16 V I	2 M	19 G S T	35 Z E	35 Z E
E	22 T N	15 H O	5 L	5 L	16 V I	8 X	35 Z E	23 D R	19 F T S	2 M	15 H O	22 T N	35 Z E	8 A	10 I	22 T N	14 Y	23 D R	35 Z E	14 Y	4 B	23 D R	16 V I	19 F T S	15 H O
F	22 T N	35 Z E	23 D R	19 F T S	2 K	35 Z E	23 D R	35 Z E	35 Z E	3 J	10 S A	8 X	10 I	10 S A	19 G S T	35 Z E	14 Y	19 G S T	8 A	16 V I	19 F T S	10 S A	5 L	19 G S T	35 Z E
G	23 D R	22 T N	15 H O	15 H O	22 T N	1 C	23 D R	35 Z E	2 R	10 S A	16 V I	22 T N	14 Y	35 Z E	23 D R	3 O	35 Z E	19 F T S	19 F T S	15 H O	22 T N	35 Z E	8 A	10 I	22 T N
H	14 Y	23 D R	35 Z E	14 Y	19 G S T	8 A	16 V I	23 D R	19 G S T	35 Z E	35 Z E	22 T N	2 K	15 H O	10 I	22 T N	14 Y	35 Z E	14 Y	10 S A	23 D R	35 Z E	19 G S T	15 H O	5 U
J	35 Z E	19 F T S	35 Z E	22 T N	19 G S T	5 U	5 P	19 G S T	23 D R	10 I	8 X	1 W	24 G S T	15 H O	8 X	8 A	10 S A	2 R	5 U	35 Z E	23 D R	19 F T S	5 U	10 I	23 D R
K	3 E	19 G S T	15 H O	22 T N	16 V I	3 E	8 A	19 G S T	8 X	8 X															

d. No impossible sequences are brought to light, and, moreover, several long words, nearly complete, stand out in the text. Note the following portions:

```
     A21
     H B Z D S G V H T F
(1)  O ? E R A S I O N T
               T       S

     C15
     T V T Z B D V F H T Z D F
(2)  N I N E ? R I T O N E R T
                   S         S

     F22
     S L G Z D T H H T
(3)  A ? S E R N O O N
         T
```

The words are obviously OPERATIONS, NINE PRISONERS, and AFTERNOON. The value G_e si clearly T_p; that of F_e is S_p; and the following additional values are certain:

$$B_e=P_p \qquad L_e=F_p$$

31. Completing the solution.—*a*. Each time an additional value is obtained, substitution is at once made throughout the cryptogram. This leads to the determination of further values, in an ever-widening circle, until all the identifications are firmly and finally established, and the message is completely solved. In this case the decipherment is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	S	F	D	Z	F	I	O	G	H	L	P	Z	F	G	Z	D	Y	S	P	F	H	B	Z	D	S
	A	S	R	E	S	U	L	T	O	F	Y	E	S	T	E	R	D	A	Y	S	O	P	E	R	A
B	G	V	H	T	F	U	P	L	V	D	F	G	Y	V	J	V	F	V	H	T	G	A	D	Z	Z
	T	I	O	N	S	B	Y	F	I	R	S	T	D	I	V	I	S	I	O	N	T	H	R	E	E
C	A	I	T	Y	D	Z	Y	F	Z	J	Z	T	G	P	T	V	T	Z	B	D	V	F	H	T	Z
	H	U	N	D	R	E	D	S	E	V	E	N	T	Y	N	I	N	E	P	R	I	S	O	N	E
D	D	F	X	S	B	G	I	D	Z	Y	V	T	X	O	I	Y	V	T	E	F	V	M	G	Z	Z
	R	S	C	A	P	T	U	R	E	D	I	N	C	L	U	D	I	N	G	S	I	X	T	E	E
E	T	H	L	L	V	X	Z	D	F	M	H	T	Z	A	I	T	Y	D	Z	Y	B	D	V	F	H
	N	O	F	F	I	C	E	R	S	X	O	N	E	H	U	N	D	R	E	D	P	R	I	S	O
F	T	Z	D	F	K	Z	D	Z	Z	J	S	X	I	S	G	Z	Y	G	A	V	F	S	L	G	Z
	N	E	R	S	W	E	R	E	E	V	A	C	U	A	T	E	D	T	H	I	S	A	F	T	E
G	D	T	H	H	T	C	D	Z	R	S	V	T	Y	Z	D	O	Z	F	F	H	T	Z	A	I	T
	R	N	O	O	N	Q	R	E	M	A	I	N	D	E	R	L	E	S	S	O	N	E	H	U	N
H	Y	D	Z	Y	G	A	V	D	G	Z	Z	T	K	H	I	T	Y	Z	Y	S	D	Z	G	H	U
	D	R	E	D	T	H	I	R	T	E	E	N	W	O	U	N	D	E	D	A	R	E	T	O	B
J	Z	F	Z	T	G	U	P	G	D	I	X	W	G	H	X	A	S	R	U	Z	D	F	U	I	D
	E	S	E	N	T	B	Y	T	R	U	C	K	T	O	C	H	A	M	B	E	R	S	B	U	R
K	E	G	H	T	V	E	A	G	X	X															
	G	T	O	N	I	G	H	T	X	X															

Message: AS RESULT OF YESTERDAYS OPERATIONS BY FIRST DIVISION THREE HUNDRED SEVENTY NINE PRISONERS CAPTURED INCLUDING SIXTEEN OFFICERS ONE HUNDRED PRISONERS WERE EVACUATED THIS AFTERNOON REMAINDER LESS ONE HUNDRED THIRTEEN WOUNDED ARE TO BE SENT BY TRUCK TO CHAMBERSBURG TONIGHT

b. The solution should, as a rule, not be considered complete until an attempt has been made to discover all the elements underlying the general system and the specific key to a message. In this case, there is no need to delve further into the general system, for it is merely one of monoalphabetic substitution with a mixed cipher alphabet. It is necessary or advisable, however, to reconstruct the cipher alphabet because this may give clues that later may become valuable.

c. Cipher alphabets should, as a rule, be reconstructed by the cryptanalyst in the form of *enciphering* alphabets because they will then usually be in the form in which the encipherer used them. This is important for two reasons. First, if the sequence in the cipher component gives evidence of system in its construction or if it yields clues pointing toward its derivation from a keyword or a key-phrase, this may often corroborate the identifications already made and may lead directly to additional identifications. A word or two of explanation is advisable here. For example, refer to the skeletonized enciphering alphabet given at the end of par. 29*b:*

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S				Z				V					T	H			D	G	F	I					
																			F	G						

Suppose the cryptanalyst, looking at the sequence DGFI or DFGI in the cipher component, suspects the presence of a keyword-mixed alphabet. Then DFGI is certainly a more plausible sequence than DGFI. Again, noting the sequence S . . . Z . . . V . . . . TH . . D, he might have an idea that the keyword begins after the Z and that the TH is followed by AB or BC. This would mean that either P, Q_p=A, B_c or B, C_c. Assuming that P, Q_p=A, B_c, he refers to the frequency distribution and finds that the assumptions P_p=A_c and Q_p=B_c are not good; on the other hand, assuming that P, Q_p=B, C_c, the frequency distribution gives excellent corroboration. A trial of these values would materially hasten solution because it is often the case in cryptanalysis that if the value of a very low-frequency letter can be surely established it will yield clues to other values very quickly. Thus, if Q_p is definitely identified it almost invariably will identify U_p, and will give clues to the letter following the U_p, since it must be a vowel. In the case under discussion the identification PQ_p=BC_c would have turned out to be correct. For the foregoing reason an attempt should always be made in the early stages of the analysis to determine, if possible, the basis of construction or derivation of the cipher alphabet; as a rule this can be done only by means of the enciphering alphabet, and not the deciphering alphabet. For example, the skeletonized *deciphering* alphabet corresponding to the enciphering alphabet directly above is as follows:

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain				R		T	S	O	U										A	N		I				E
						S	T																			

Here no evidences of a keyword-mixed alphabet are seen at all. However, if the enciphering alphabet has been examined and shows no evidences of systematic construction, the deciphering alphabet should then be examined with this in view, because occasionally it is the deciphering alphabet which shows the presence of a key or keying element, or which has been systematically derived from a word or phrase. The second reason why it is important to try to discover the basis of construction or derivation of the cipher alphabet is that it affords clues to the general type of keywords or keying elements employed by the enemy. This is a psychological factor, of course, and may be of assistance in subsequent studies of his traffic. It merely gives a clue to the general type of thinking indulged in by certain of his cryptographers.

d. In the case of the foregoing solution, the complete enciphering alphabet is found to be as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	

Obviously, the letter Q, which is the only letter not appearing in the cryptogram, should follow P in the cipher component. Note now that the latter is based upon the keyword LEAVENWORTH, and that this particular cipher alphabet has been composed by shifting the mixed sequence based upon this keyword five intervals to the right so that the key for the message is $A_p=S_c$. Note also that the deciphering alphabet fails to give any evidence of keyword construction based upon the word LEAVENWORTH.

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	H	P	Q	R	G	S	T	O	U	V	W	F	X	J	L	Y	Z	M	A	N	B	I	K	C	D	E

e. If neither the enciphering or the deciphering alphabet exhibits characteristics which give indication of derivation from a keyword by some form of mixing or disarrangement, the latter is nevertheless not finally excluded as a possibility. The student is referred to Section IX of *Elementary Military Cryptography*, wherein will be found methods for deriving mixed alphabets by transposition methods applied to keyword-mixed alphabets. For the reconstruction of such mixed alphabets the cryptanalyst must use ingenuity and a knowledge of the more common methods of suppressing the appearance of keywords in the mixed alphabets.

32. **General notes on the foregoing solution.**—*a.* The example solved above is admittedly a more or less artificial illustration of the steps in analysis, made so in order to demonstrate general principles. It was easy to solve because the frequencies of the various cipher letters corresponded quite well with the normal or expected frequencies. However, all cryptograms of the same monoalphabetical nature can be solved along the same general lines, after more or less experimentation, depending upon the length of the cryptogram, the skill, and the experience of the cryptanalyst.

b. It is no cause for discouragement if the student's initial attempts to solve a cryptogram of this type require much more time and effort than were apparently required in solving the foregoing purely illustrative example. It is indeed rarely the case that *every* assumption made by the cryptanalyst proves in the end to have been correct; more often is it the case that a good many of his initial assumptions are incorrect, and that he loses much time in casting out the erroneous ones. The speed and facility with which this elimination process is conducted is in many cases all that distinguishes the expert from the novice.

c. Nor will the student always find that the initial classification into vowels and consonants can be accomplished as easily and quickly as was apparently the case in the illustrative example. The principles indicated are very general in their nature and applicability, and there are, in addition, some other principles that may be brought to bear in case of difficulty. Of these, perhaps the most useful are the following:

(1) In normal English it is unusual to find two or three consonants in succession, each of high frequency. If in a cryptogram a succession of three or four letters of high-frequency appear in succession, it is practically certain that at least one of these represents a vowel.[3]

---

[3] Sequences of seven consonants are not impossible, however, as in STRENGTH THROUGH.

(2) Successions of three vowels are rather unusual in English.[4] Practically the only time this happens is when a word ends in two vowels and the next word begins with a vowel.[5]

(3) When two letters already classified as vowel-equivalents are separated by a sequence of six or more letters, it is either the case that one of the supposed vowel-equivalents is incorrect, or else that one or more of the intermediate letters is a vowel-equivalent.[6]

(4) Reference to Table 7–B of Appendix 1 discloses the following:

Distribution of first 18 digraphs forming 25 percent of English text

Number of consonant-consonant digraphs	4
Number of consonant-vowel digraphs	6
Number of vowel-consonant digraphs	8
Number of vowel-vowel digraphs	0

Distribution of first 53 digraphs forming 50 percent of English text

Number of consonant-consonant digraphs	8
Number of consonant-vowel digraphs	23
Number of vowel-consonant digraphs	18
Number of vowel-vowel digraphs	4

The latter tabulation shows that of the first 53 digraphs which form 50 percent of English text, 41 of them, that is, over 75 percent, are combinations of a vowel with a consonant. In short, in normal English the vowels and the high-frequency consonants are in the long run distributed fairly evenly and regularly throughout the text.

(5) As a rule, repetitions of trigraphs in the cipher text are composed of high-frequency letters forming high-frequency combinations. The latter practically always contain at least one vowel; in fact, if reference is made to Table 10–A of Appendix 1, it will be noted that 36 of the 56 trigraphs having a frequency of 100 or more contain one vowel, 17 of them contain two vowels, and only three of them contain no vowel. In the case of tetragraph repetitions, Table 11–A of Appendix 1 shows that no tetragraph listed therein fails to contain at least one vowel; 27 of them contain one vowel, 25 contain two vowels, and 2 contain three vowels.

(6) Quite frequently when two known vowel-equivalents are separated by six or more letters none of which seems to be of sufficiently high frequency to represent one of the vowels A E I O, the chances are good that the cipher-equivalent of the vowel U or Y is present.

(7) The letter Q is invariably followed by U; the letters J and V are invariably followed by a vowel.

d. In the foregoing example the amount of experimentation or "cutting and fitting" was practically nil. (This is not true of real cases as a rule.) Where such experimentation is necessary, the underscoring of all repetitions of several letters is very essential, as it calls attention to peculiarities of structure that often yield clues.

---

[4] Note that the word RADIOED, past tense of the verb RADIO, is coming into usage.

[5] A sequence of seven vowels is not impossible, however, as in THE WAY YOU EARN.

[6] Some cryptanalysts place a good deal of emphasis upon this principle as a method of locating the remaining vowels after the first two or three have been located. They recommend that the latter be underlined throughout the text and then all sequences of five or more letters showing no underlines be studied attentively. Certain letters which occur in several such sequences are sure to be vowels. An arithmetical aid in the study is as follows: Take a letter thought to be a good possibility as the cipher equivalent of a vowel (hereafter termed a *possible vowel-equivalent*) and find the length of each interval from the possible vowel-equivalent to the next *known* (fairly surely determined) vowel-equivalent. Multiply the interval by the number of times this interval is found. Add the products and divide by the total number of intervals considered. This will give the *mean* interval for that possible vowel-equivalent. Do the same for all the other possible vowel-equivalents. The one for which the mean is the greatest is most probably a vowel-equivalent. Underline this letter throughout the text and repeat the process for locating additional vowel-equivalents, if any remain to be located.

e. After a few basic assumptions of values have been made, if short words or skeletons of words do not become manifest, it is necessary to make further assumptions for unidentified letters. This is accomplished most often by assuming a word.[7] Now there are two places in every message which lend themselves more readily to successful attack by the assumption of words than do any other places—the very beginning and the very end of the message. The reason is quite obvious, for although words may begin or end with almost any letter of the alphabet, they usually begin and end with but a few very common digraphs and trigraphs. Very often the association of letters in peculiar combinations will enable the student to note where one word ends and the next begins. For example suppose, E, N, S, and T have been definitely identified, and a sequence like the following is found in a cryptogram:

. . . E N T S N E . . .

Obviously the break between two words should fall either after the S of E N T S or after the T of E N T, so that two possibilities are offered: . . . E N T S / N E . . ., or . . . E N T / S N E . . .. Since in English there are very few words with the initial trigraph S N E, it is most likely that the proper division is . . . E N T S / N E . . .. Obviously, when several word divisions have been found, the solution is more readily achieved because of the greater ease with which assumptions of additional new values may be made.

33. The "probable word" method; its value and applicability.—*a*. In practically all cryptanalytic studies, short-cuts can often be made by assuming the presence of certain words in the message under study. Some writers attach so much value to this kind of an "attack from the rear" that they practically elevate it to the position of a method and call it the "intuitive method" or the "probable-word method." It is, of course, merely a refinement of what in every-day language is called "assuming" or "guessing" a word in the message. The value of making a "good guess" can hardly be overestimated, and the cryptanalyst should never feel that he is accomplishing a solution by an illegitimate subterfuge when he has made a fortunate guess leading to solution. A correct assumption as to plain text will often save hours or days of labor, and sometimes there is no alternative but to try to "guess a word", for occasionally a system is encountered the solution of which is absolutely dependent upon this artifice.

b. The expression "good guess" is used advisedly. For it is "good" in two respects. First, the cryptanalyst must use care in making his assumptions as to plain-text words. In this he must be guided by extraneous circumstances leading to the assumption of *probable* words—not just any words that come to his mind. Therefore he must use his imagination but he must nevertheless carefully control it by the exercise of *good* judgment. Second, only if the "guess" is correct and leads to solution, or at least puts him on the road to solution, is it a *good* guess. But, while realizing the usefulness and the time and labor-saving features of a solution by assuming a probable word, the cryptanalyst should exercise discretion in regard to how long he may continue in his efforts with this method. Sometimes he may actually waste time by adhering to the method too long, if straightforward, methodical analysis will yield results more quickly.

c. Obviously, the "probable-word" method has much more applicability when working upon material the general nature of which is known, than when working upon more or less isolated communications exchanged between correspondents concerning whom or whose activities nothing is known. For in the latter case there is little or nothing that the imagination can seize upon as a background or basis for the assumptions.[8]

---

[7] This process does not involve anything more mysterious than ordinary, logical reasoning; there is nothing of the subnormal or supernormal about it. If cryptanalytic success seems to require processes akin to those of medieval magic, if "hocus-pocus" is much to the fore, the student should begin to look for items that the claimant of such success has carefully hidden from view, for the mystification of the uninitiated. If the student were to adopt as his personal motto for all his cryptanalytic ventures the quotation (from Tennyson's poem *Columbus*) appearing on the back of the title page of this text, he will frequently find "short cuts" to his destination and will not too often be led astray!

d. Very frequently, the choice of probable words is aided or limited by the number and positions of repeated letters. These repetitions may be *patent*—that is, externally visible in the cryptographic text as it originally stands—or they may be *latent*—that is, externally invisible but susceptible of being made patent as a result of the analysis. For example, in a monoalphabetic substitution cipher, such as that discussed in the preceding paragraph, the repeated letters are directly exhibited in the cryptogram; later the student will encounter many cases in which the repetitions are latent, but are made patent by the analytical process. When the repetitions are patent, then the *pattern* or *formula* to which the repeated letters conform is of direct use in assuming plain-text words; and when the text is in word-lengths, the pattern is obviously of even greater assistance. Suppose the cryptanalyst is dealing with military text, in which case he may expect such words as DIVISION, BATTALION, etc., to be present in the text. The positions of the repeated letter I in DIVISION, of the reversible digraph AT, TA in BATTALION, and so on, constitute for the experienced cryptanalyst tell-tale indications of the presence of these words, even when the text is not divided up into its original word lengths.

e. The important aid that a study of word patterns can afford in cryptanalysis warrants the use of definite terminology and the establishment of certain data having a bearing thereon. The phenomenon herein under discussion, namely, that many words are of such construction as regards the number and positions of repeated letters as to make them readily identifiable, will be termed *idiomorphism* (from the Greek "idios"=one's own, individual, peculiar+"morphe"=form). Words which show this phenomenon will be termed *idiomorphic.* It will be useful to deal with the idiomorphisms symbolically and systematically as described below.

f. When dealing with cryptograms in which the word lengths are determined or specifically shown, it is convenient to indicate their lengths and their repeated letters in some easily recognized manner or by formulas. This is exemplified, in the case of the word DIVISION, by the formula ABCBDBEF; in the case of the word BATTALION, by the formula ABCCBDEFG. If the cryptanalyst, during the course of his studies, makes note of striking formulas he has encountered, with the words which fit them, after some time he will have assembled a quite valuable body of data. And after more or less complete lists of such formulas have been established in some systematic arrangement, a rapid comparison of the idiomorphs in a specific cryptogram with those in his lists will be feasible and will often lead to the assumption of the correct word. Such lists can be arranged according to word length, as shown herewith:

3/aba : DID, EVE, EYE.
abb : ADD, ALL, ILL, OFF, etc.
4/abac : ARAB, AWAY, etc.
abca : AREA, BOMB, DEAD, etc.
abbc : . . .
abcb : . . .
etc. etc.

---

[8] General Givierge in his *Cours de Cryptographie* (p. 121) says: "However, expert cryptanalysts often employ such details as are cited above [in connection with assuming the presence of 'probable words'], and the experience of the years 1914 to 1918, to cite only those, prove that in practice one often has at his disposal elements of this nature, permitting assumptions much more audacious than those which served for the analysis of the last example. The reader would therefore be wrong in imagining that such fortuitous elements are encountered only in cryptographic works where the author deciphers a document that he himself enciphered. Cryptographic correspondence, if it is extensive, and if sufficiently numerous working data are at hand, often furnishes elements so complete that an author would not dare use all of them in solving a problem for fear of being accused of obvious exaggeration."

g. When dealing with cryptographic text in which the lengths of the words are not indicated or otherwise determinable, lists of the foregoing nature are not so useful as lists in which the words (or parts of words) are arranged according to the intervals between identical letters, in the following manner:

1 Interval	2 Intervals	3 Intervals	Repeated digraphs
-DiD-	AbbAcy	AbeyAnce	COCOa
-EvE-	ArAbiA	hAbitAble	dERER
-EyE-	AbiAtive	lAborAtory	ICICle
dIvIsion	AboArd	AbreAst	ININg
revIsIon	-AciA-	AbroAd	bAGgAGe
etc.	etc.	etc.	etc.

34. Solution of additional cryptograms produced by the same cipher component.—*a.* To return, after a rather long digression, to the cryptogram solved in pars. 28–31, once the cipher component of a cipher alphabet has been reconstructed, subsequent messages which have been enciphered by means of the same cipher component may be solved very readily, and without recourse to the principles of frequency, or application of the probable-word method. It has been seen that the illustrative cryptogram treated in paragraphs 24–31 was enciphered by juxtaposing the cipher component against the normal sequence so that $A_p=S_c$. It is obvious that the cipher component may be set against the plain component at any one of 26 different points of coincidence, each yielding a different cipher alphabet. After a cipher component has been reconstructed, however, it becomes a *known* sequence, and the method of converting the cipher letters into their plain-component equivalents and then completing the plain-component sequence begun by each equivalent can be applied to solve any cryptogram which has been enciphered by that cipher component.

b. An example will serve to make the process clear. Suppose the following message, passing between the same two stations as before, was intercepted shortly after the first message had been solved:

IYEWK CERNW OFOSE LFOOH EAZXX

It is assumed that the same cipher component was used, but with a different key letter. First the initial two groups are converted into their plain-component equivalents by setting the cipher component against the normal sequence at any arbitrary point of coincidence. The initial letter of the former may as well be set against A of the latter, with the following result:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z

Cryptogram	I	Y	E	W	K	C	E	R	N	W	. . .
Equivalents	P	Y	B	F	R	L	B	H	E	F	. . .

The normal sequence initiated by each of these conversion equivalents is now completed, with the results shown in Fig. 15. Note the plain-text generatrix, CLOSEYOURS, which manifests itself without further analysis. The rest of the message may be read either by continuing the same process, or, what is even more simple, the key letter of the message may now be determined quite readily and the message deciphered by its means.

I	Y	E	W	K	C	E	R	N	W
P	Y	B	F	R	L	B	H	E	F
Q	Z	C	G	S	M	C	I	F	G
R	A	D	H	T	N	D	J	G	H
S	B	E	I	U	O	E	K	H	I
T	C	F	J	V	P	F	L	I	J
U	D	G	K	W	Q	G	M	J	K
V	E	H	L	X	R	H	N	K	L
W	F	I	M	Y	S	I	O	L	M
X	G	J	N	Z	T	J	P	M	N
Y	H	K	O	A	U	K	Q	N	O
Z	I	L	P	B	V	L	R	O	P
A	J	M	Q	C	W	M	S	P	Q
B	K	N	R	D	X	N	T	Q	R
*C	L	O	S	E	Y	O	U	R	S
D	M	P	T	F	Z	P	V	S	T
E	N	Q	U	G	A	Q	W	T	U
F	O	R	V	H	B	R	X	U	V
G	P	S	W	I	C	S	Y	V	W
H	Q	T	X	J	D	T	Z	W	X
I	R	U	Y	K	E	U	A	X	Y
J	S	V	Z	L	F	V	B	Y	Z
K	T	W	A	M	G	W	C	Z	A
L	U	X	B	N	H	X	D	A	B
M	V	Y	C	O	I	Y	E	B	C
N	W	Z	D	P	J	Z	F	C	D
O	X	A	E	Q	K	A	G	D	E

c. In order that the student may understand without question just what is involved in the latter step, that is, discovering the key letter after the first two or three groups have been deciphered by the conversion-completion process, the foregoing example will be used. It was noted that the first cipher group was finally deciphered as follows:

Cipher	I Y E W K
Plain	C L O S E

Now set the cipher component against the normal sequence so that $C_p=I_c$. Thus:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D

It is seen here that when $C_p=I_c$ then $A_p=F_c$. This is the key for the entire message. The decipherment may be completed by direct reference to the foregoing cipher alphabet. Thus:

Cipher	I Y E W K	C E R N W	O F O S E	L F O O H	E A Z X X
Plain	C L O S E	Y O U R S	T A T I O	N A T T W	O P M X X

Message: CLOSE YOUR STATION AT TWO PM

d. The student should make sure that he understands the fundamental principles involved in this quick solution, for they are among the most important principles in cryptanalytics. How useful they are will become clear as he progresses into more and more complex cryptanalytic studies.

SECTION VII

MULTILITERAL SUBSTITUTION WITH SINGLE-EQUIVALENT CIPHER ALPHABETS

Paragraph
Analysis of multiliteral, monoalphabetic substitution systems........ 35
Historically interesting examples........ 36

35. Analysis of multiliteral, monoalphabetic substitution systems.—*a*. Substitution methods in general may be classified into uniliteral and multiliteral systems.[1] In the former there is a strict "one-to-one" correspondence between the length of the units of the plain and those of the cipher text; that is, each letter of the plain text is replaced by a single character in the cipher text. In the latter this correspondence is no longer $1_p{:}1_c$ but may be $1_p{:}2_c$, where each letter of the plain text is replaced by a combination of two characters in the cipher text; or $1_p{:}3_c$, where a 3-character combination in the cipher text represents a single letter of the plain text, and so on. A cipher in which the correspondence of the $1_p{:}1_c$ type is termed uniliteral in character; one in which it is of the $1_p{:}2_c$ type, biliteral; $1_p{:}3_c$, triliteral, and so on. Those beyond the $1_p{:}1_c$ type are classed together as *multiliteral*.

b. When a multiliteral system employs biliteral equivalents, the cipher alphabet is said to be bipartite. Such alphabets are composed of a set of 25 or 26 combinations of a limited number of characters taken in pairs. An example of such an alphabet is the following.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M
Cipher	WW	WH	WI	WT	WE	HW	HH	HI	HT	HT	HE	IW	IH
Plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	II	IT	IE	TW	TH	TI	TT	TE	EW	EH	EI	ET	EE

The foregoing alphabet is derived from the cipher square, or *matrix*, shown in Fig. 15.

(2)

		W	H	I	T	E
	W	A	B	C	D	E
	H	F	G	H	I–J	K
(1)	I	L	M	N	O	P
	T	Q	R	S	T	U
	E	V	W	X	Y	Z

FIGURE 15.

c. If a message is enciphered by means of the foregoing bipartite alphabet the cryptogram is still monoalphabetic in character. A frequency distribution based upon pairs of letters will obviously have all the characteristics of a simple, uniliteral distribution for a monoalphabetic substitution cipher.

---

[1] See Sec. VII, *Advanced Military Cryptography.*

d. Ciphers of this type, as well as of those of the multiliteral (triliteral, quadraliteral, . . .) type are readily detected externally by virtue of the fact that the cryptographic text is composed of but a very limited number of different characters. They are handled in exactly the same manner as are uniliteral, monoalphabetic substitution ciphers. So long as the same character, or combination of characters, is always used to represent the same plain-text letter, and so long as a given letter of the plain text is always represented by the same character or combination of characters, the substitution is strictly monoalphabetic and can be handled in the simple manner described under Par. 31 of this text.

e. An interesting example in which the cipher equivalents are quinqueliteral groups and yet the resulting cipher is strictly monoalphabetic in character is found in the cipher system invented by Sir Francis Bacon over 300 years ago. Despite its antiquity the system possesses certain features of merit which are well worth noting. Bacon[2] proposed the following cipher alphabet, composed of permutations of two elements taken five at a time:[3]

A=aaaaa
B=aaaab
C=aaaba
D=aaabb
E=aabaa
F=aabab
G=aabba
H=aabbb
I–J=abaaa
K=abaab
L=ababa
M=ababb
N=abbaa
O=abbab
P=abbba
Q=abbbb
R=baaaa
S=baaab
T=baaba
U–V=baabb
W=babaa
X=babab
Y=babba
Z=babbb

If this were all there were to Bacon's invention it would be hardly worth bringing to attention. But what he pointed out, with great clarity and simple examples, was how such an alphabet might be used to convey a secret message by enfolding it in an innocent, external message which might easily evade the strictest kind of censorship. As a very crude example, suppose that a message is written in capital and lower case letters, any capital letter standing for an "a" element of the cipher alphabet, and any small letter, for a "b" element. Then the external sentence "All is well with me today" can be made to contain the secret message "Help." Thus:

$$\underbrace{\begin{matrix}\text{A L l} & \text{i s}\\ \text{a a b} & \text{b b}\end{matrix}}_{\text{H}}\quad \underbrace{\begin{matrix}\text{W E l L} & \text{W}\\ \text{a a b a} & \text{a}\end{matrix}}_{\text{E}}\quad \underbrace{\begin{matrix}\text{I t H} & \text{m E}\\ \text{a b a} & \text{b a}\end{matrix}}_{\text{L}}\quad \underbrace{\begin{matrix}\text{T o d a Y}\\ \text{a b b b a}\end{matrix}}_{\text{P}}$$

Instead of employing such an obvious device as capital and small letters, suppose that an "a" element be indicated by a very slight shading, or a very slightly heavier stroke. Then a secret message might easily be thus enfolded within an external message of exactly opposite meaning. The number of possible variations of this basic scheme is very high. The fact that the characters of the cryptographic text are hidden in some manner or other has, however, no effect upon the strict monoalphabeticity of the scheme.

---

[2] For a true picture of this cipher, the explanation of which is often distorted beyond recognition even by cryptographers, see Bacon's own description of it as contained in his *De Augmentis Scientiarum* (*The Advancement of Learning*), as translated by any first-class editor, such as Gilbert Watts (1640) or Ellis, Spedding, and Heath (1857, 1870). The student is cautioned, however, not to accept as true any alleged "decipherments" obtained by the application of Bacon's cipher to literary works of the 16th century. These readings are purely subjective.

[3] In the 16th Century, the letters I and J were used interchangeably, as were also U and V. Bacon's alphabet was called by him a "biliteral alphabet" because it employs permutations of two letters. But from the cryptanalytic standpoint the significant point is that each plain-text letter is represented by a 5-character equivalent. Hence, present terminology requires that this alphabet be referred to as a *quinqueliteral alphabet*.

36. Historically interesting examples.—*a.* Two examples of historical interest will be cited in this connection as illustrations. During the campaign for the presidential election of 1876 many cipher messages were exchanged between the Tilden managers and their agents in several states where the voting was hotly contested. Two years later the New York Tribune[4] exposed many irregularities in the campaign by publishing the decipherments of many of these messages. These decipherments were achieved by two investigators employed by the Tribune, and the plain text of the messages seems to show that illegal attempts and measures to carry the election for Tilden were made by his managers. Here is one of the messages:

JACKSONVILLE, Nov. 16 (1876).

GEO. F. RANEY, Tallahassee.

P p y y e m n s n y y y p i m a s h n s y y s s i t e p a a e n s h n s
p e n s s h n s m m p i y y s n p p y e a a p i e i s s y e s h a i n s s s p
e e i y y s h n y n s s s y e p i a a n y i t n s s h y y s p y y p i n s y y
s s i t e m e i p i m m e i s s e i y y e i s s i t e i e p y y p e e i a a s s
i m a a y e s p n s y y i a n s s s e i s s m m p p n s p i n s s n p i n s i m
i m y y i t e m y y s s p e y y m m n s y y s s i t s p y y p e e p p p m a
a a y y p i i t
L'Engle goes up tomorrow.

DANIEL.

Examination of the message discloses that only ten different letters are used. It is probable, therefore, that what one has here is a cipher which employs a bipartite alphabet and in which combinations of two letters represent single letters of the plain text. The message is therefore rewritten in pairs and substitution of arbitrary letters for the pairs is made, as seen below:

PP	YY	EM	NS	NY	YY	PI	MA	SH	NS	YY	SS	etc.
A	B	C	D	E	B	F	G	H	D	B	I	etc.

A triliteral frequency distribution is then made and analysis of the message along the lines illustrated in the preceding section of this text yields solution, as follows:

JACKSONVILLE, *Nov. 16.*

GEO. F. RANEY, *Tallahassee:*

Have Marble and Coyle telegraph for influential men from Delaware and Virginia. Indications of weakening here. Press advantage and watch Board. L'Engle goes up tomorrow.

DANIEL.

b. The other example, using numbers, is as follows:

JACKSONVILLE, Nov. 17.

S. PASCO and E. M. L'ENGLE:

84 55 84 25 93 34 82 31 31 75 93 82 77 33 55 42
93 20 93 66 77 66 33 84 66 31 31 93 20 82 33 66
52 48 44 55 42 82 48 89 42 93 31 82 66 75 31 93

DANIEL.

---

[4] New York Tribune, Extra No. 44, *The Cipher Dispatches*, New York, 1879.

There were, of course, several messages of like nature, and examination disclosed that only 26 different numbers in all were used. Solution of these ciphers followed very easily, the decipherment of the one given above being as follows:

JACKSONVILLE, *Nov. 17.*

S. PASCO and E. M. L'ENGLE:

Cocke will be ignored, Eagan called in. Authority reliable.

DANIEL.

c. The Tribune experts gave the following alphabets as the result of their decipherments:

AA=O
AI=U
EI=I
EM=V
EN=Y
EP=C
IA=K
IM=S
IT=D
MA=B
MM=G
NN=J
NS=E
NY=M
PE=T
PI=R
PP=H
SH=L
SN=P
SP=W
SS=N
YE=F
YI=X
YY=A

20=D
25=K
27=S
31=L
33=N
34=W
39=P
42=R
44=H
48=T
52=U
55=O
62=X
66=A
68=F
75=B
77=G
82=I
84=C
87=V
89=Y
93=E
96=M
99=J

They did not attempt to correlate these alphabets, or at least they say nothing about a possible relationship. The present author has, however, reconstructed the rectangle upon which these alphabets are based, and it is given below (fig. 16).

2d Letter or Number

1st Letter or Number	H 1	I 2	S 3	P 4	A 5	Y 6	M 7	E 8	N 9	T 0
H 1										
I 2					K		S			D
S 3	L		N	W					P	
P 4		R		H				T		
A 5		U			O					
Y 6		X				A		F		
M 7					B		G			
E 8		I		C			V		Y	
N 9			E			M			J	
T 0										

FIGURE 16.

It is amusing to note that the conspirators selected as their key a phrase quite in keeping with their attempted illegalities—HIS PAYMENT—for bribery seems to have played a considerable part in that campaign. The blank squares in the diagram probably contained proper names, numbers, etc.

SECTION VIII

MULTILITERAL SUBSTITUTION WITH MULTIPLE-EQUIVALENT CIPHER ALPHABETS

Paragraph
Purpose of providing multiple-equivalent cipher alphabets........ 37
Solution of a simple example........ 38
Solution of more complicated example........ 39
A subterfuge to prevent decomposition of cipher text into component units........ 40

37. Purpose of providing multiple-equivalent cipher alphabets.—*a*. It has been seen that the characteristic frequencies of letters composing normal plain text, the associations they form in combining to form words, and the peculiarities certain of them manifest in such text all afford direct clues by means of which ordinary monoalphabetic substitution encipherments of such plain text may be more or less speedily solved. This has led to the introduction of simple methods for disguising or suppressing the manifestations of monoalphabeticity, so far as possible. Basically these methods are multiliteral and they will now be presented.

b. Multiliteral substitution may be of two types: (1) That wherein each letter of the plain text is represented by one and only one multiliteral equivalent. For example, in the Francis Bacon cipher described in Par. 35*e*, the letter K_p is invariably represented by the permutation abaab. For this reason this type of system may be more completely described as *monoalphabetic, multiliteral substitution with single-equivalent cipher alphabets.*

(2) That wherein, because of the large number of equivalents made available by the combinations and permutations of a limited number of elements, each letter of the plain text may be represented by several multiliteral equivalents which may be selected at random. For example, if 3-letter combinations are employed there are available 26^3 or 17,576 equivalents for the 26 letters of the plain text; they may be assigned in equal numbers of different equivalents for the 26 letters, in which case each letter would be representable by 676 different 3-letter equivalents; or they may be assigned on some other basis, for example, proportionately to the relative frequencies of plain-text letters. For this reason this type of system may be more completely described as a *monoalphabetic, multiliteral substitution with a multiple-equivalent cipher alphabet*. Some authors term such a system "simple substitution with multiple equivalents"; others term it *monoalphabetic substitution with variants*. For the sake of brevity, the latter designation will be employed in this text.[1]

c. The primary object of monoalphabetic substitution with variants is, as has been mentioned above, to provide several values which may be employed at random in a simple substitution of cipher equivalents for the plain-text letters. In this connection, reference is made to Section X of *Elementary Military Cryptography*, wherein several of the most common methods for producing and using variants are set forth.

---

[1] My attention has been called to the contradiction in terminology involved in the designation "monoalphabetic substitution with variants", and when judged by my own definition of the term, as given in *Elementary Military Cryptography*, it must be admitted that the criticism is warranted. Systems employing *complete* cipher alphabets which form the basis for the choice of variant values are, strictly speaking, not monoalphabetic but polyalphabetic in nature. But what shall be said of those systems in which there are no complete cipher alphabets for the selection of alternative or variant values but only a few variants for the high-frequency letters? I recognize, of course, that ease or difficulty in the solution of specific types of systems should not be the determining factor in a systematic nomenclature in cryptography. The equation $x^2=16$ is not designated a linear equation because it is easy to solve; it is designated a quadratic equation and although it is easier to solve than $x^2+x=16$ it is a quadratic nevertheless. The analogy is, of course, clear: a system which is technically polyalphabetic should not be designated monoalphabetic because it is usually easier to solve than a polyalphabetic system. Indeed, there are systems of monoalphabetic substitution with variants which are more difficult to solve than certain types of polyalphabetic systems, for example, those using several standard alphabets in a cyclic manner. However, the designation "monoalphabetic substitution with variants" has become so firmly established in the literature and is so descriptive of the actual process followed in encipherment that I hesitate to change it at this date in favor of some less descriptive but more accurate designation, such as "multiliteral substitution with multiple-equivalent cipher alphabets", although I *have* used it as the title of this section.

d. A word or two concerning the underlying theory from the cryptanalytic point of view of monoalphabetic substitution with variants, may not be amiss. Whereas in simple or single-equivalent, monoalphabetic substitution it is seen that—

(1) The same letter of the plain text is invariably represented by but one and always the same character of the cryptogram, and

(2) The same character of the cryptogram invariably represents one and always the same letter of the plain text;

In multiliteral substitution with multiple equivalents (monoalphabetic substitution with variants) it is seen that—

(1) The same letter of the plain text may be represented by one or more different characters of the cryptogram, but

(2) The same character of the cryptogram nevertheless invariably represents one and always the same letter of the plain text.

38. *Solution of a simple example.*—*a.* The following cryptogram has been enciphered by a set of four alphabets similar to the following:

A	B	C	D	E	F	G	H	I-J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	01	02	03	04	05	06	07
35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	26	27	28	29	30	31	32	33	34
68	69	70	71	72	73	74	75	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
87	88	89	90	91	92	93	94	95	96	97	98	99	00	76	77	78	79	80	81	82	83	84	85	86

The keyword here is TRIP[1]. In enciphering a message the equivalents are to be selected at random from among the four variants for each letter. The steps in solving a message produced by such a scheme will now be scrutinized.

CRYPTOGRAM

68321 09022 48057 65111 88648 42036 45235 09144 05764 22684
00225 57003 97357 14074 82524 40768 51058 93074 92188 47264
09328 04255 06186 79882 85144 45886 32574 55136 56019 45722
76844 68350 45219 71649 90528 65106 11886 44044 89669 70553
18491 06985 48579 33684 50957 70612 09795 29148 56109 08546
62062 65509 32800 32568 97216 44282 34031 84989 68564 53789
12530 77401 68494 38544 11368 87616 56905 20710 58864 67472
22490 09136 62851 24551 35180 14230 50886 44084 06231 12876
05579 58980 29503 99713 32720 36433 82689 04516 52263 21175
06445 72255 68951 86957 76095 67215 53049 08567 9730

b. Assuming that the foregoing remarks had not been made and that the cryptogram has just been submitted for solution with no information concerning it, the first step is to make a preliminary study to determine whether the cryptogram involves cipher or code. The cryptogram appears in 5-figure groups, which may indicate either cipher or code. A few remarks will be made at this point with reference to the method of determining whether a cryptogram composed of figure groups is in code or cipher, using the foregoing example.

c. In the first place, if the cryptogram contains an even number of digits, as for example 494 in the foregoing message, this leaves open the possibility that it may be cipher, composed of 247 pairs of digits; were the number of digits an exact *odd* multiple of five, such as 125, 135, etc., the possibility that the cryptogram is in code of the 5-figure group type must be considered. Next, a preliminary study is made to see if there are many repetitions, and what their characteristics are. If the cryptogram is code of the 5-figure group type, then such repetitions as appear should *generally* be in whole groups of five digits, and they should be visible in the text just as the message stands, unless the code message has undergone encipherment also. If the cryptogram is in cipher, then the repetitions should extend beyond the 5-digit groupings; if they conform to any definite groupings at all they should for the most part contain even numbers of digits since each letter is probably represented by a pair of digits. If no clues of the foregoing nature are present, doubts will be dissolved by making a detailed study of frequencies.

---

[1] The letter corresponding to the lowest number in each line of the diagram showing the cipher alphabets is a key letter. Thus, in the 1st line 01=T; in the 2d line 26=R; etc.

d. A simple 4-part frequency distribution is therefore decided upon. Shall the alphabet be assumed to be a 25- or a 26-character one? If the former, then the 2-digit pairs from 01 to 00 fall into exactly four groups each corresponding to an alphabet. Since this is the most common scheme of drawing up such alphabets, let it be assumed to be true of the present case. The following distributions result from the breaking up of the text into 2-digit pairs.

01—///
02—
03—////
04—/
05—𝍸
06—𝍸 /
07—///
08—
09—////
10—////
11—𝍸
12—///
13—/
14—/
15—/
16—///
17—
18—𝍸 /
19—
20—/
21—//
22—𝍸
23—//
24—
25—/

26—///
27—
28—/
29—/
30—///
31—
32—𝍸 /
33—/
34—/
35—//
36—𝍸
37—/
38—
39—/
40—///
41—
42—////
43—/
44—𝍸 /
45—𝍸 /
46—///
47—
48—///
49—𝍸
50—𝍸

51—𝍸
52—𝍸
53—///
54—
55—////
56—𝍸
57—𝍸 /
58—//
59—
60—
61—
62—//
63—
64—𝍸 /
65—
66—/
67—//
68—𝍸 //
69—//
70—/
71—/
72—////
73—
74—////
75—/

76—𝍸 /
77—/
78—
79—/
80—///
81—
82—////
83—/
84—𝍸 /
85—𝍸 /
86—///
87—
88—////
89—𝍸
90—𝍸 /
91—///
92—/
93—/
94—/
95—///
96—
97—𝍸 /
98—/
99—
00—//

e. If the student will bring to bear upon this problem the principles he learned in Section V of this text, he will soon realize that what he now has before him are four, simple, monoalphabetic frequency distributions similar to those involved in a monoalphabetic substitution cipher using standard cipher alphabets. The realization of this fact immediately provides the clue to the next step: "fitting each of the distributions to the normal." (See Par. 17*b*). This can be done without difficulty in this case (remembering that a 25-letter alphabet is involved and assuming that I and J are the same letter) and the following alphabets result:

01—I-J
02—K
03—L
04—M
05—N
06—O
07—P
08—Q
09—R
10—S
11—T
12—U
13—V
14—W
15—X
16—Y
17—Z
18—A
19—B
20—C
21—D
22—E
23—F
24—G
25—H

26—U
27—V
28—W
29—X
30—Y
31—Z
32—A
33—B
34—C
35—D
36—E
37—F
38—G
39—H
40—I-J
41—K
42—L
43—M
44—N
45—O
46—P
47—Q
48—R
49—S
50—T

51—N
52—O
53—P
54—Q
55—R
56—S
57—T
58—U
59—V
60—W
61—X
62—Y
63—Z
64—A
65—B
66—C
67—D
68—E
69—F
70—G
71—H
72—I-J
73—K
74—L
75—M

76—E
77—F
78—G
79—H
80—I-J
81—K
82—L
83—M
84—N
85—O
86—P
87—Q
88—R
89—S
90—T
91—U
92—V
93—W
94—X
95—Y
96—Z
97—A
98—B
99—C
00—D

f. The keyword is seen to be JUNE and the first few groups of the cryptogram decipher as follows:

68	32	10	90	22	48	05	76	51	11	88	64	84	20	36	45	23
E	A	S	T	E	R	N	E	N	T	R	A	N	C	E	O	F

g. From the detailed procedure given above, the student should be able to draw his own conclusions as to the procedure to be followed in solving cryptograms produced by methods which are more or less simple variations of that just discussed. In this connection he is referred to Section X of *Elementary Military Cryptography*, wherein a few of these variations are mentioned.

h. Possibly the most important of the variations is that in which a rectangle such as that shown in Fig. 17 is employed.

	1	2	3	4	5	6	7	8	9	0
1, 4, 7	A	B	C	D	E	F	G	H	I	J
2, 5, 8	K	L	M	N	O	P	Q	R	S	T
3, 6, 9	U	V	W	X	Y	Z	-	,	:	;

FIGURE 17

In the solution of cases of this kind, repetitions would play their usual role, with the modifications noted below in Par. 39. Once an entering wedge has been forced, through the identification of one or more repeated words such as BATTALION, DIVISION, etc., the entire enciphering matrix would soon be reconstructed. It may be added that the frequency distribution for the text of a single long message or several short ones enciphered by such a system would show characteristic phenomena, the most important of which are, first, that the distribution for a matrix such as shown in Fig. 17 would practically follow the normal and, second, that the distribution for the 2d digit of pairs would show more marked crests and troughs than the distribution for the 1st digit. For example, the initial digits 1, 4, and 7 (for the numbers 10–19, 40–49, and 70–79, inclusive) would apply to the distribution for the letters A to J, inclusive; the initial digits 2, 5, and 8 would apply to the distribution for the letters K to T, inclusive. The total weighted frequency values for these two groups of letters are about equal. Therefore, the frequencies of the initial digits 1, 2, 4, 5, 7, and 8 would be approximately equal. But consider the final digit 5 in the numbers 15, 45, 75, 25, 55, and 85; its total frequency is composed of the sum of the frequencies of E_p, O_p, and Y_p, two of which are high-frequency letters; whereas in the case of the final digit 6, its total frequency is composed of the sum of the frequencies of F_p, P_p, and Z_p, all of which are low-frequency letters. The two cases would show a marked difference in frequency, in the proportion of $1319+844+208=2371$ to $205+243+6=454$, or about 5 to 1. Of course, the letters may be inserted within the enciphering matrix in a keyword-mixed or even in a random order; the numbers may be applied to the rectangle in a random order. But these variations, while increasing the difficulty in solution, by no means make the latter as great as may be thought by the novice.

39. Solution of a more complicated example.—*a.* As soon as a beginner in cryptography realizes the consequences of the fact that letters are used with greatly varying frequencies in normal plain text, a brilliant idea very speedily comes to him. Why not disguise the natural frequencies of letters by a system of substitution using many equivalents, and let the numbers of equivalents assigned to the various letters be more or less in direct proportion to the normal frequencies of the letters? Let E, for example, have 13 or more equivalents; T, 10; N, 9; etc., and thus (he thinks) the enemy cryptanalyst can have nothing in the way of tell-tale or characteristic frequencies to use as an entering wedge.

b. If the text available for study is small in amount and if the variant values are wholly independent of one another, the problem can become exceedingly difficult. But in practical military communications such methods are rarely encountered, *because the volume of text is usually great enough to permit of the establishment of equivalent values.* To illustrate what is meant, suppose a set of cryptograms produced by the monoalphabetic-variant method described above shows the following two sets of groupings in the text:

SET A	SET B
12-37-02-79-68-13-03-37-77	71-12-02-51-23-05-77
82-69-02-79-13-68-23-37-35	11-82-51-02-03-05-35
82-69-51-16-13-13-78-05-35	11-91-02-02-23-37-35
91-05-02-01-68-42-78-37-77	97-12-51-02-78-69-77

An examination of these groupings would lead to the following tentative conclusions with regard to probable equivalents:

12, 82, 91
05, 37, 69
02, and 51
01, 16, 79
13, 42, 68
03, 23, 78
35, and 77

The establishment of these equivalencies would sooner or later lead to the finding of additional sets of equal values. The completeness with which this can be accomplished will determine the ease or difficulty of solution. Of course, if many equivalencies can be established the problem can then be reduced practically to monoalphabetic terms and a speedy solution can be attained.

c. Theoretically, the determination of equivalencies may seem to be quite an easy matter, but practically it may be very difficult, because the cryptanalyst can never be *certain* that a combination showing what may appear to be a variant value is really such, and is not a different word. For example, take the groups—

17–82–31–82–14–63, and
27–82–40–82–14–63

Here one might suspect that 17 and 27 represent the same letter, 31 and 40 another letter. But it happens that one group represents the word MANAGE, the other DAMAGE.

d. When reversible combinations are used as variants, the problem is perhaps a bit more simple. For example, using the accompanying Fig. 18 for encipherment, two messages with the same initial words, REFERENCE YOUR, may be enciphered as follows:

	K,Z	Q,V	B,H	M,R	D,L
W,S	N	H	A	O	E
F,X	D	T	M	F	P
G,J	Q	B	U	I	V
C,N	G	X	R	C	S
P,T	Z	L	Y	W	K

FIGURE 18.

```
      R   E    F    E   R    E   N    C   E   Y    O   U    R
(1)  N H W D R  X L S H C  D W W Z N  R S L H P  S R B J C  H
(2)  C H D W R  X S L H N  D W Z W N  R L S H P  R W J B N  H
```

The experienced cryptanalyst, noting the appearance of the very first few groups, assumes that he is here confronted with a case involving biliteral reversible equivalents, with variants.

e. The probable-word method of solution may be used, but with a slight variation introduced by virtue of the fact that, regardless of the system, *letters of low frequency in plain text remain infrequent.* Hence, suppose a word containing low-frequency letters, but in itself a rather common word strikingly idiomorphic in character is sought as a "probable word"; for example, words such as $\underline{\text{C}}\text{A}\underline{\text{V}}\text{ALR}\underline{\text{Y}}$, $\text{ATTAC}\underline{\text{K}}$, and $\underline{\text{P}}\text{RE}\underline{\text{P}}\text{ARE}$. Writing such a word on a slip of paper, it is slid one interval at a time under the text, which has been marked so that the high and low-frequency characters are indicated. Each coincidence of a low-frequency letter of the text with a low-frequency letter of the assumed word is examined carefully to see whether the adjacent text letters correspond in frequency with the other letters of the assumed word; or, if the latter presents repetitions, whether there are correspondences between repetitions in the text and those in the word. Many trials are necessary but this method will produce results when the difficulties are otherwise too much for the cryptanalyst to overcome.

40. A subterfuge to prevent decomposition of cipher text into component units.—*a.* A few words should be added with regard to certain subterfuges which are sometimes encountered in monoalphabetic substitution with variants, and which, if not recognized in time, cause considerable delays. These have to deal with the insertion of nulls so as to prevent the cryptanalyst from breaking up the text into its real cryptographic units. The student should take careful note of the last phrase; the mere insertion of symbols having the same characteristics as the symbols of the cryptographic text, except that they have no meaning, is not what is meant. This class of nulls rarely achieves the purpose for which they are intended. What is really meant can best be explained in connection with an example. Suppose that a 5 x 5 checkerboard design with the row and column indicators shown in Fig. 19 is adopted for encipherment. Normally, the cipher units would consist of 2-letter combinations of the indicators, invariably giving the row indicator first (by agreement).

	V	G	I	W	D
	A	H	P	S	M
	T	O	E	B	N
	F	U	R	L	C
V,A,T,F	A	B	C	D	E
G,H,O,U	F	G	H	I–J	K
I,P,E,R	L	M	N	O	P
W,S,B,L	Q	R	S	T	U
D,M,N,C	V	W	X	Y	Z

FIGURE 19.

The phrase COMMANDER OF SPECIAL TROOPS might be enciphered thus:

C	O	M	M	A	N	D	E	R	O	F	. . .
VI	EB	PH	IU	FT	IE	AB	TM	WO	PW	GT	. . .

These would normally then be arranged in 5-letter groups, thus:

VIEBP HIUFT IEABT MWOPW GT . . .

b. It will be noted, however, that only 20 of the 26 letters of the alphabet have been employed as row and column indicators, leaving J, K, Q, X, Y, and Z unused. Now, suppose these five letters are used as nulls, *not in pairs, but as individual letters inserted at random* just before the real text is arranged in 5-letter groups. Occasionally, a pair of nulls is inserted. Thus, for example:

VIEXB PHKIU FJXTI EAJBT MWOQP WGKTY

The cryptanalyst, after some study, suspecting a biliteral cipher, proceeds to break up the text into pairs:

VI EX BP HK IU FJ XT IE AJ BT MW OQ PW GK TY

Compare this set of 2-letter combinations with the correct set. Only 4 of the 15 pairs are "proper" units. It is easy to see that without a knowledge of the *existence* of the nulls, and even with a knowledge, if he does not know *which* letters are nulls, the cryptanalyst would be confronted with a problem for the solution of which a fairly large amount of text might be necessary. The careful employment of the variants also very materially adds to the security of the method because repetitions can be rather effectively suppressed.

c. From the cryptographic standpoint, the fact that in this system the cryptographic text is more than twice as long as the plain text constitutes a serious disadvantage. From the cryptanalytic standpoint, the masking of the cipher units constitutes the most important source of strength of the system; this, coupled with the use of variants, makes it a bit more difficult system to solve, despite its monoalphabeticity.

SECTION IX

POLYGRAPHIC SUBSTITUTION SYSTEMS

	Paragraph
Monographic and polygraphic substitution systems	41
Tests for identifying digraphic substitution	42
General procedure in the analysis of digraphic substitution ciphers	43
Analysis of digraphic substitution ciphers based upon 4-square checkerboard matrices	44
Analysis of ciphers based upon other types of checkerboard matrices	45
Analysis of the Playfair cipher system	46

41. Monographic and polygraphic substitution systems.—*a.* The student is now referred to Sections VII and VIII of *Advanced Military Cryptography*, wherein polygraphic systems of substitution are discussed from the cryptographic point of view. These will now be discussed from the cryptanalytic point of view.

b. Although the essential differences between polyliteral and polygraphic substitution are treated with some detail in Section VII of *Advanced Military Cryptography*, a few additional words on the subject may not be amiss at this point.

c. The two primary divisions of substitution systems into (1) uniliteral and multiliteral methods and into (2) monographic and polygraphic methods are both based upon considerations as to the *number of elements* constituting the plain-text and the equivalent cipher-text units. In uniliteral as well as in monographic substitution, each plain-text unit consists of a single element and each cipher-text unit consists of a single element. The two terms uniliteral and monographic are therefore identical in significance, as defined cryptographically. It is when the terms multiliteral and polygraphic are examined that an essential difference is seen. In multiliteral substitution the plain-text unit always consists of a single element (one letter) and the cipher-text unit consists of a group of two or more elements; when biliteral, it is a pair of elements, when triliteral, it is a set of three elements, and so on. In what will herein be designated as true or complete polygraphic substitution the plain-text unit consists of two or more elements forming an *indivisible compound;* the cipher-text unit usually consists of a corresponding number of elements.[1] When the number of elements comprising the plain-text units is fixed and always two, the system is *digraphic;* when it is three, the system is *trigraphic;* when it is four, *tetragraphic;* and so on.[2] It is important to note that in true or complete polygraphic substitution the elements combine to form indivisible compounds having properties different from those of either of the constituent letters. For example, in uniliteral substitution AB_p may yield XY_c and AC_p may yield XZ_c; but in true digraphic substitution $\overline{AB}_p$ may yield $\overline{XY}_c$ and $\overline{AC}_p$ may yield $\overline{QN}_c$. A difference in identity of one letter affects the whole result.[3] An analogy is found in chemistry, when two elements combine to form a molecule, the latter usually having properties quite different from those of either of the constituent elements. For example: sodium, a metal, and chlorine, a gas, combine to form sodium chloride, common table salt. Furthermore, sodium and fluorine, also a gas similar in many respects to chlorine, combine to form sodium fluoride, which is much different from table salt. Partial and pseudo-polygraphic substitution will be treated under subparagraphs *d* and *e* below.

---

[1] The qualifying adverb "usually" is employed because this correspondence is not essential. For example, if one should draw up a set of 676 arbitrary single signs, it would be possible to represent the 2-letter pairs from AA to ZZ by single symbols. This would still be a digraphic system.

[2] In this sense a code system is merely a polygraphic substitution system in which the number of elements constituting the plain-text units is variable.

[3] For this reason the two letters are marked by a ligature, that is, by a bar across their tops.

d. Another way of looking at polygraphic substitution is to regard the elements comprising the plain-text units as being enciphered individually and polyalphabetically by a fairly large number of separate alphabets. For example, in a digraphic system in which 676 pairs of plain-text letters are representable by 676 cipher-text pairs assigned at random, this is equivalent to having a set of 26 different alphabets for enciphering one member of the pairs, and another set of 26 different alphabets for enciphering the other member of the pairs. According to this viewpoint the different alphabets are brought into play by the particular combination of letters forming each plain-text pair. This is, of course, quite different from systems wherein the various alphabets are brought into play by more definite rules; it is perhaps this very absence of definite rules guiding the selection of alphabets which constitutes the cryptographic strength of this type of polygraphic system.

e. When regarded in the light of the preceding remarks, certain systems which at first glance seem to be polygraphic, in that groupings of plain-text letters are treated as units, on closer inspection are seen to be only partially polygraphic, or pseudo-polygraphic in character. For example, in a system in which encipherment is by pairs and yet one of the letters in each pair is enciphered monoalphabetically, the other letter, polyalphabetically, the method is only *psuedo*-polygraphic. Cases of this type are shown in Section VII of *Advanced Military Cryptography.* Again, in a system in which encipherment is by pairs and the encipherments of the left-hand and right-hand members of the pairs show group relationships, this is not pseudo-polygraphic but only *partially* polygraphic. Cases of this type are also shown in the text referred to above.

f. The fundamental purpose of polygraphic substitution is again the suppression of the frequency characteristics of plain text, just as is the case in monoalphabetic substitution with variants; but here this is accomplished by a different method, the latter arising from a somewhat different approach to the problem involved in producing cryptographic security. When the substitution involves replacement of *single* letters in a monoalphabetic system, the cryptogram can be solved rather readily. Basically the reason for this is that the principles of frequency and the laws of probability, applied to individual units of the text (single letters), have a very good opportunity to manifest themselves. A given volume of text of say n plain-text letters, enciphered purely monoalphabetically, affords n cipher characters, and the same number of cipher units. The same volume of text, enciphered digraphically, still affords n cipher characters but only $\frac{n}{2}$ cipher units. Statistically speaking, the sample within which the laws of probability now apply has been cut in half. Furthermore, from the point of view of frequency, the very noticeable diversity in the frequencies of individual letters, leading to the marked crests and troughs of the uniliteral frequency distribution, is no longer so strikingly in evidence in the frequencies of digraphs. Therefore, although true digraphic encipherment, for example, cuts the cryptographic textual units in half, the difficulty of solution is not doubled, but, if a matter of judgment arising from practical experience can be expressed or approximated mathematically, squared or cubed.

g. Sections VII and VIII of *Advanced Military Cryptography* show various methods for the derivation of polygraphic equivalents and for handling these equivalents in cryptographing and decryptographing messages. The most practicable of those methods are digraphic in character and for this reason their solution will be treated in a somewhat more detailed manner than will trigraphic methods. The latter can be passed over with the simple statement that their analysis requires much text to permit of solution by the frequency method, and hard labor. Fortunately, they are infrequently encountered because they are difficult to manipulate without extensive tables.[4] If the latter are required they must be compiled in the form of a book or pamphlet. If one is willing to go that far, one might as well include in such document more or less extensive lists of words and phrases, in which case the system falls under the category of code and not cipher.

42. Tests for identifying digraphic substitution.—*a*. The tests which are applied to determine whether a given cryptogram is digraphic in character are usually rather simple. If there are many repetitions in the cryptogram and yet the uniliteral-frequency distribution gives no clear-cut indications of monoalphabeticity; if most of the repetitions contain an even number of letters; and if the cryptogram contains an even number of letters, it may be assumed to be digraphic in nature.

b. The student should first try to determine whether the substitution is completely digraphic, or only partially digraphic, or pseudo-digraphic in character. As mentioned above, there are cases in which, although the substitution is effected by taking pairs of letters, one of the members of the pairs is enciphered monoalphabetically, the other member, polyalphabetically. A distribution based upon the letters in the odd positions and one based upon those in the even positions should be made. If one of these is clearly monoalphabetic, then this is evidence that the message represents a case of pseudo-digraphism of the type here described. By attacking the monoalphabetic portion of the messages, solution can soon be reached by slight variation of the usual method, the polyalphabetic portion being solved by the aid of the context and considerations based upon the probable nature of the substitution chart. (See Tables 2, 3, and 4 of *Advanced Military Cryptography.*) It will be noted that the charts referred to show definite symmetry in their construction.

c. On the other hand, if the foregoing steps prove fruitless, it may be assumed that the cryptogram is completely digraphic in character.

d. Just as certain statistical tests may be applied to a cryptogram to establish its monoalphabeticity, so also may a statistical test be applied to a cryptogram for the purpose of establishing its digraphicity. The nature of this test and its method of application will be discussed in a subsequent text.

43. General procedure in the analysis of digraphic substitution ciphers.—*a*. The analysis of cryptograms which have been produced by digraphic substitution is accomplished largely by the application of the simple principles of frequency of digraphs, with the additional aid of such special circumstances as may be known to or suspected by the cryptanalyst. The latter refer to peculiarities which may be the result of the particular method employed in obtaining the equivalents of the plain-text digraphs in the cryptographing process. In general, however, only if there is sufficient text to disclose the normal phenomena of repetition will solution be feasible or possible.

b. However, when a digraphic system is employed in regular service, there is little doubt but that traffic will rapidly accumulate to an amount more than sufficient to permit of solution by simple principles of frequency. Sometimes only two or three long messages, or a half dozen of average length are sufficient. For with the identification of only a few cipher digraphs, larger portions of messages may be read because the skeletons of words formed from the few high-frequency digraphs very definitely limit the values that can be inserted for the intervening unidentified digraphs. For example, suppose that the plain-text digraphs TH, ER, IN, IS, OF, NT, and TO have been identified by frequency considerations, corroborated by a tentatively identified long repetition; and suppose also that the enemy is known to be using a quadricular table of 676 cells containing digraphs showing reciprocal equivalence between plain and cipher-text digraphs. Suppose the message begins as follows (in which the assumed values have been inserted):

---

[4] A patent has been granted upon a rather ingenious machine for automatically accomplishing true polygraphic substitution, but it has not been placed upon the market. See U. S. Patent No. 1845947 issued in 1932 to Weisner and Hill. In U. S. Patent No. 1515680 issued to Henkels in 1924, there is described a mechanism which also produces polygraphic substitution.

XQ	VO	ZI	LK	AP	OL	ZX	PV	QN	IK	OL	UK	AL	HN	LK	VL
FO		TH	IN		NT		RE			NT	NO			IN	
BN	OZ	KU	DY	EL	LE	YW									
SI		ON	TO												

The words FOURTH INFANTRY REGIMENT are readily recognized. The reciprocal pairs EL_c and LE_c suggest ATTACK. The beginning of the message is now completely disclosed: FOURTH INFANTRY REGIMENT NOT YET IN POSITION TO ATTACK. The values more or less automatically determined are $VO_c=UR_p$, $AL_c=TY_p$, $HN_c=ET_p$, $VL_c=PO_p$, $OZ_c=TI_p$, $YW_c=CK_p$.

c. Once a good start has been made and a few words have been solved, subsequent work is quite simple and straightforward. A knowledge of enemy correspondence, including data regarding its most common words and phrases, is of great assistance in breaking down new digraphic tables of the same nature but with different equivalents.

d. The foregoing remarks also apply to the details of solution in cases of partially digraphic substitution.

44. Analysis of digraphic substutution ciphers based upon 4-square checkerboard matrices.—*a.* In Section VIII of *Advanced Military Cryptography* there are shown various examples of digraphic substitution based upon the use of checkerboard designs. These may be considered cases of partially digraphic substitution, in that in the checkerboard system there are certain relationships between plain-text digraphs having common elements and their corresponding cipher-text digraphs, which will also have common elements. For example, take the following 4-square checkerboard matrix:

	B	W	G	R	M	O	P	A	U	L	
	N	Y	V	X	E	H	Z	Q	D	F	
1	S	I	C	T	K	K	I	T	S	C	3
	U	P	L	A	O	M	W	R	B	G	
	D	Z	F	Q	H	E	Y	X	N	V	
	W	A	L	E	S	C	X	K	P	B	
	F	H	U	I	T	O	M	Y	D	V	
4	P	X	B	K	C	S	A	E	W	L	2
	N	Z	R	Q	G	G	Z	Q	N	R	
	D	M	V	Y	O	T	H	I	F	U	

FIGURE 20.

Here $BC_p=OW_c$, $BO_p=OF_c$, $BS_p=OP_c$, $BG_p=ON_c$ and $BT_p=OD_c$. In each case when B_p is the initial letter of the plain-text pair, the initial letter of the cipher-text equivalent is O_c. This, of course, is the direct result of the method; it means that the encipherment is monoalphabetic for the first half of each of these *five* plain-text pairs, polyalphabetic for the second half. This relationship holds true for *four* other groups of pairs beginning with B_p. In other words, there are five alphabets employed, not 25. Thus, this case differs from the case discussed under Par. 42*b* only in that the monoalphabeticity is not complete for one-half of all the pairs, but only among the members of certain groups of pairs. In a completely digraphic system using a 676-cell randomized matrix, such relationships are entirely absent and for this reason the system is cryptographically more secure than the checkerboard system.

b. From the foregoing, it is clear that when solution has progressed sufficiently to disclose a few values, the insertion of letters within the cells of the checkerboard matrix to give the plain-text and cipher relationships indicated by the solved values immediately leads to the disclosure of additional values. Thus, the solution of only a few values soons leads to the breakdown of the entire matrix.

c. (1) The following example will serve to illustrate the procedure. Let the message be as follows:

```
    1 2 3 4 5   6 7 8 9 10   11 12 13 14 15   16 17 18 19 20   21 22 23 24 25   26 27 28 29 30
A.  H F C A P   G O Q I L    B  S  P  K  M    N  D  U  K  E    O  H  Q  N  F    B  O  R  U  N
B.  Q C L C H   Q B Q B F    H  M  A  F  X    S  I  O  K  O    Q  Y  F  N  S    X  M  C  G  Y
C.  X I F B E   X A F D X    L  P  M  X  H    H  R  G  K  G    Q  K  Q  M  L    F  E  Q  Q  I
D.  G O I H M   U E O R D    C  L  T  U  F    E  Q  Q  C  G    Q  N  H  F  X    I  F  B  E  X
E.  F L B U Q   F C H Q O    Q  M  A  F  T    X  S  Y  C  B    E  P  F  N  B    S  P  K  N  U
F.  Q I T X E   U Q M L F    E  Q  Q  I  G    O  I  E  U  E    H  P  I  A  N    Y  T  F  L  B
G.  F E E P I   D H P C G    N  Q  I  H  B    F  H  M  H  F    X  C  K  U  P    D  G  Q  P  N
H.  C B C Q L   Q P N F N    P  N  I  T  O    R  T  E  N  C    O  B  C  N  T    F  H  H  A  Y
I.  Z L Q C I   A A I Q U    C  H  T  P  C    B  I  F  G  W    K  F  C  Q  S    L  Q  M  C  B
J.  O Y C R Q   Q D P R X    F  N  Q  M  L    F  I  D  G  C    C  G  I  O  G    O  I  H  H  F
K.  I R C G G   G N D L N    O  Z  T  F  G    E  E  R  R  P    I  F  H  O  T    F  H  H  A  Y
L.  Z L Q C I   A A I Q U    C  H  T  P
```

(2) The cipher having been tested for standard alphabets (by the method of completing the normal components) and found to give negative results, a uniliteral-frequency distribution is made. It is as follows:

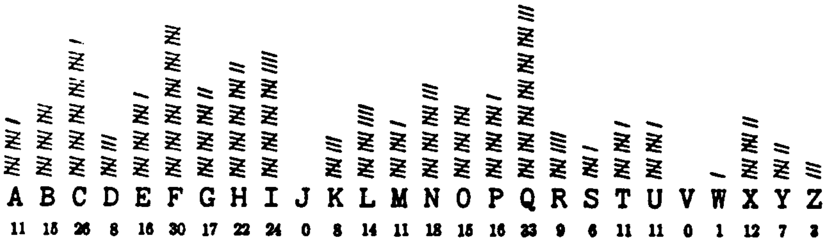


FIGURE 21.

(3) At first glance this may appear to the untrained eye to be a monoalphabetic frequency distribution but upon closer inspection it is noted that aside from the frequencies of four or five

letters the frequencies for the remaining letters are not very dissimilar. There are, in reality, no very marked crests and troughs, certainly not as many as would be expected in a monoalphabetic substitution cipher of equal length.

(4) The message having been carefully examined for repetitions of 4 or more letters, all of them are listed:

	Frequency	Located in lines
TFHHAYZLQCIAAIQUCHTP (20 letters)	2	H and K.
QMLFEQQIGOI (11 letters)	2	C and F.
XIFBEX (6 letters)	2	C and D.
FEQQ	3	C, D, F.
QMLF	3	C, F, J.
BFHM	2	B and G.
BSPK	2	A and E.
GOIH	2	D and J.

Since there are quite a few repetitions, two of considerable length, since all but one of them contain an even number of letters, and since the message also contains an even number of letters, 344, digraphic substitution is suspected. The cryptogram is transcribed in 2-letter groups, for greater convenience in study. It is as follows:

Message transcribed in pairs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B.	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C.	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM	LF	EQ	QI
D.	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E.	FL	BU	QF	CH	QO	QM	AF	TX	SY	CB	EP	FN	BS	PK	NU
F.	QI	TX	EU	QM	LF	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G.	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN
H.	CB	CQ	LQ	PN	FN	PN	IT	OR	TE	NC	CB	CN	TF	HH	AY
J.	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM	CB
K.	OY	CR	QQ	DP	RX	FN	QM	LF	ID	GC	CG	IO	GO	IH	HF
L.	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M.	ZL	QC	IA	AI	QU	CH	TP								

It is noted that all the repetitions listed above break up properly into digraphs except in one case, viz, FEQQ in lines C, D, and F. This seems rather strange, and at first thought one might suppose that a letter was dropped out or was added in the vicinity of the FEQQ in line D. But it is immediately seen that the FE QQ in line D has no relation at all to the .F EQ Q. in lines C and F, and that the F EQ Q in line D is merely an accidental repetition.

(5) A digraphic frequency distribution[5] is made and is shown in Fig. 22.

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A						///			//															//	
B						//										/		//		/					
C	/	𝍸					////	///			/		/			//	/								
D															/					/			/		
E														/	//	//	/			/			//		
F		///			//						/		////												
G			/		/		/							///		/						/		/	
H						////		///				//		/	//	/									
I	///			//	/	//		///			/			//			/		/						
K					/	/	/							/						/					
L		/	/			///							/		/	/									
M			/										/							/			/		
N			/	/												/				/				/	
O								/								/	//							/	/
P				/			/			//			///												
Q			///			/			///	/		𝍸 /	//	/		//				//				/	
R				/			/								/								/		
S											/												/	/	
T					/	////									//					/			//		
U					/								/												
V																									
W																									
X			/						//									/							
Y																									
Z											//														

FIGURE 22.

(6) The appearance of the foregoing distribution for this message is quite characteristic of that for a digraphic substitution cipher. There are many blank cells; although there are many cases in which a digraph appears only once, there are quite a few in which a digraph appears two or three times, four cases in which a digraph appears four times, and two cases in which a digraph appears five times. The absence of the letter J is also noted; this is often the case in a digraphic system based upon a checkerboard matrix.

---

[5] The distinction between "digraphic" and "biliteral" is based upon the following consideration. In a biliteral (or diliteral) distribution every two successive letters of the text would be grouped together to form a pair. For example, a biliteral distribution of ABCDEF would tabulate the pairs AB, BC, CD, DE, and EF. In a digraphic distribution only successive pairs of the text are tabulated. For example, ABCDEF would yield only AB, CD, and EF.

(7) In another common type of checkerboard system known as the Playfair cipher, described in Par. 46, one of the telltale indications besides the absence of the letter J is the absence of double letters, that is, two successive identical letters. The occurrence of the double letters GG, HH, and QQ in the message under investigation eliminates the possibility of its being a Playfair cipher. The simplest thing to assume is that a 4-square checkerboard is involved. One with normal alphabets in Sections 1 and 2 is therefore set down (Fig. 23*a*).

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P						3
	Q	R	S	T	U						
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I–J	K	
4						L	M	N	O	P	2
						Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*a*.

(8) The recurrence of the group QMLF, three times, and at intervals suggesting that it might be a sentence separator, leads to the assumption that it is the word STOP. The letters Q, M, L, and F are therefore inserted in the appropriate cells in Sections 3 and 4 of the diagram. Thus (Fig. 23*b*):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U				Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*b*.

These placements seem logical. Moreover, in Section 3 the number of cells between L and Q is just one less than enough to contain all the letters M to P, inclusive, and suggests that either N or O is in the keyword portion of the sequence, that is, near the top of Section 3. Without making a commitment in the matter, suppose both N and O, for the present, be inserted in the cell between M and P. Thus (Fig. 23c):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23c.

(9) Now, if the placement of P in Section 3 is correct, the cipher equivalent of TH_p will be $P\theta_e$, and there should be a group of adequate frequency to correspond. Noting that PN_e occurs three times, it is assumed to be TH_p and the letter N is inserted in the appropriate cell in Section 4. Thus (Fig. 23d):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23d.

(10) It is about time to try out these assumed values in the message. The proper insertions are made, with the following results:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B.	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C.	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM ST	LF OP	EQ	QI →
D.	← GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E.	FL	BU	QF	CH	QO	QM ST	AF	TX	SY	CB	EP	FN	BS	PK	NU
F.	QI	TX	EU	QM ST	LF OP	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G.	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN TH
H.	CB	CQ	LQ	PN TH	FN	PN TH	IT	OR	TE	NC	CB	CN	TF	HH	AY →
J.	← ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM ST	CB
K.	OY	CR	QQ	DP	RX	FN	QM ST	LF OP	ID	GC	CG	IO	GO	IH	HF
L.	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY →
M.	← ZL	QC	IA	AI	QU	CH	TP								

(11) So far no impossible combinations are in evidence. Beginning with group H4 in the message is seen the following sequence:

P N F N P N
T H . . T H

Assume it to be THAT THE. Then $AT_p=FN_e$, and the letter N is to be inserted in row 4 column 1. But this is inconsistent with previous assumptions, since N in Section 4 has already been tentatively placed in row 2 column 4 of Section 4. Other assumptions for FN_e are made: that it is, IS_p (THIS TH...); that it is EN_p (THEN TH...); but the same inconsistency is apparent. In fact the student will see that FN_e must represent a digraph ending in F, G, H, I–J, or K, since N_e is tentatively located on the same line as these letters in Section 2. Now FN_e occurs 4 times in the message. The digraph it represents *must* be one of the following:

DF, DG, DH, DI, DJ, DK
IF, IG, IH, II, IJ, IK
JF, JG, JH, JI, JJ, JK
OF, OG, OH, OI, OJ, OK
TK,
YF, YG, YH, YI, YJ, YK

Of these the only one likely to be repeated 4 times is OF, yielding THOFTH which may be
PNFNPN
a part of

.NORTHOFTHE. or .SOUTHOFTHE.
CQLQPNFNPNIT CQLQPNFNPNIT

In either case, the position of the F in Section 3 is excellent: F . . . L in row 3. There are 3 cells intervening between F and L, into which G, H, I–J, and K may be inserted. It is not nearly so likely that G, H, and K are in the keyword as that I should be in it. Let it be assumed that this is the case, and let the letters be placed in the appropriate cells in Section 3. Thus (Fig. 23*e*):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

Figure 23*e*.

Let the resultant derived values be checked against the frequency distribution. If the position of H in Section 3 is correct, then the digraph ON_p, normally of high frequency should be represented several times by HF_c. Reference to Fig. 22 shows a frequency of 4 times. And HM_c, with 2 occurrences, represents NS_p. There is no need to go through all the possible corroborations.

(12) Going back to the assumption that TH..TH
PNFNPN
is part of the expression

.NORTHOFTHE. or .SOUTHOFTHE.,
CQLQPNFNPNIT CQLQPNFNPNIT

it is seen at once from Fig. 23*e* that the latter is apparently correct and not the former, because LQ_c equals OU_p and not OR_p. If $\theta S_p = CQ_c$, this means that the letter C of the digraph CQ_c must be placed in row 1 column 3 or row 2 column 3 of Section 3. Now the digraph CB_c occurs 5 times, CG_c, 4 times, CH_c, 3 times, CQ_c, 2 times. Let an attempt be made to deduce the exact position of C in Section 3 and the positions of B, G, and H in Section 4. Since F is already placed in Section

4, assume G and H directly follow it, and that B comes before it. How much before? Suppose a trial be made. Thus (Fig. 23*f*):

	A	B	C	D	E			C			
	F	G	H	I–J	K			C			
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4	B	B	B	F	G	L	M	N	O	P	2
	H		M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*f*.

By referring now to the frequency distribution, Fig. 22, after a very few minutes of experimentation it becomes apparent that the following is correct:

	A	B	C	D	E			C			
	F	G	H	I–J	K						
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4	B			F	G	L	M	N	O	P	2
	H		M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*g*.

(13) The identifications given by these placements are inserted in the text, and solution is very rapidly completed. The final checkerboard and deciphered text are given below.

	A	B	C	D	E	S	O	C	I	E	
	F	G	H	I–J	K	T	Y	A	B	D	
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N	P	Q	R	
	V	W	X	Y	Z	U	V	W	X	Z	
	E	X	P	U	L	A	B	C	D	E	
	S	I	O	N	A	F	G	H	I–J	K	
	B	C	D	F	G	L	M	N	O	P	
4	H	K	M	Q	R	Q	R	S	T	U	2
	T	V	W	Y	Z	V	W	X	Y	Z	

FIGURE 23A.

A. HFCAP GOQIL BSPKM NDUKE OHQNF BORUN
ONEHU NDRED FIRST FIELD ARTIL LERYF

B. QCLCH QBQBF HMAFX SIOKO QYFNS XMCGY
ROMPO SITIO NSINV ICINI TYOFB ARLOW

C. XIFBE XAFDX LPMXH HRGKG QKQML FEQQI
WILLB EINGE NERAL SUPPO RTSTO PDURI

D. GOIHM UEORD CLTUF EQQCG QNHFX IFBEX
NGATT ACKSP ECIAL ATTEN TIONW ILLBE

E. FLBUQ FCHQO QMAFT XSYCB EPFNB SPKNU
PAIDT OASSI STING ADVAN CEOFF IRSTB

F. QITXE UQMLF EQQIG OIEUE HPIAN YTFLB
RIGAD ESTOP DURIN GADVA NCEIT WILLP

G. FEEPI DHPCG NQIHB FHMHF XCKUP DGQPN
LACEC ONCEN TRATI ONSON WOODS NORTH

H. CBCQL QPNFN PNITO RTENC CBCNT FHHAY
ANDSO UTHOF THAYE RFARM ANDHI LLSIX

J. ZLQCI AAIQU CHTPC BIFGW KFCQS LQMCB
ZEROE IGHTD ASHAA NDONW OODSE ASTAN

K. OYCRQ QDPRX FNQML FIDGC CGIOG OIHHF
DWEST THERE OFSTO PCOMM ENCIN GATON

L. IRCGG GNDLN OZTFG EERRP IFHOT FHHAY
ETENP MSMOK EWILL BEUSE DONHI LLSIX

M. ZLQCI AAIQU CHTP
ZEROE IGHTD ASHA

d. (1) It is interesting to note how much simpler the matter becomes when the positions of the plain-text and cipher-text sections are reversed, or, what amounts to the same thing, when in encipherment the plain-text pairs are sought in the sections containing the mixed alphabets, and their cipher equivalents are taken from the sections containing the normal alphabets. For example, referring to Fig. 23*h*, suppose that sections 3–4 be used as the source of the plain-text pairs, and sections 1–2 as the source of the cipher-text pairs. Then $ON_p=DG_c$, $EH_p=AU_c$, etc.

(2) To solve a message enciphered in that manner, it is necessary merely to make a matrix in which all four sections are normal alphabets, and then perform two steps. First, the cipher text pairs are converted into their normal alphabet equivalents merely by "deciphering" the message with that matrix; the result of this operation yields two monoalphabets, one composed of the odd letters, the other of the even letters. The second step is to solve these two mono-alphabets.

(3) Where the same mixed alphabet is inserted in sections 3 and 4, the problem is still easier, since the letters resulting from the conversion into normal-alphabet equivalents all belong to the same, single-mixed alphabet.

45. Analysis of ciphers based upon other types of checkerboard matrices.—The solution of cryptograms enciphered by other types of checkerboard matrices is accomplished along lines very similar to those set forth in the foregoing example of the solution of a message prepared by means of a 4-square checkerboard matrix. There are, unfortunately, no means or tests which can be applied to determine in the early stages of the analysis exactly what type of design is involved in the *first* case under study. The author freely admits that the solution outlined in subparagraph *c* is quite artificial in that nothing is demonstrated in step (7) that obviously leads to or warrants the assumption that a 4-square checkerboard is involved. This point was passed over with the quite bald statement that this was "the simplest thing to assume"—and then the solution proceeds exactly as though this mere *hypothesis* has been definitely established. For example, the very first results obtained were based upon assuming that a certain 4-letter repetition represented the word STOP and *immediately inserting certain letters in appropriate cells in a 4-square checkerboard.* Several more assumptions were built on top of that and very rapid strides were made. What if it had not been a 4-square checkerboard at all? What if it had been a 2-square matrix of the type shown in Fig. 24?

M	A	N	U	F	O	S	Q	L	P
C	T	R	I	G	W	Z	Y	V	X
B	D	E	H	K	D	K	H	B	E
L	O	P	Q	S	A	F	U	M	N
V	W	X	Y	Z	T	G	I	C	R

FIGURE 24.

The only defense that can be made of what may seem to the student to be purely arbitrary procedure based upon the author's advance information or knowledge is the following: In the first place, in order to avoid making the explanation a too-long-drawn-out affair, it is necessary (and pedagogical experience warrants) that certain alternative hypotheses be passed over in silence. In the second place, it may now be added, *after* the principles and procedure have been elucidated (which at this stage is the primary object of this text) that if good results do not follow from a first hypothesis, the only thing the cryptanalyst can do is to reject that hypothesis, and formulate a second hypothesis. In actual practice he may have to reject a second, third, fourth, . . . *nth* hypothesis. In the end he may strike the right one—or he may not. There is no guaranty of success in the matter. In the third place, one of the objects of this text is to show how certain systems, if employed for military purposes, can readily be broken down. Assuming that a checkerboard system is in use, and that daily changes in keywords are made, it is possible that the traffic of the first day might give considerable difficulty in solution, if the type of checkerboard were not known to the cryptanalyst. But the second or third day's traffic would be easy to solve, because by that time the cryptanalytic personnel would have analyzed the system and thus learned what type of checkerboard the enemy is using.

46. Analysis of the Playfair cipher system.—*a.* An excellent example of a practical, partially digraphic system is the Playfair cipher.[6] It was used for a number of years as a field cipher by the British Army, before and during the World War, and for a short time, also during that war, by certain units of the American Expeditionary Forces.

b. Published solutions[7] for this cipher are quite similar basically and vary only in minor details. The earliest, that by Lieut. Mauborgne, used straightforward principles of frequency to establish the values of three or four of the most frequent digraphs. Then, on the assumption that in most cases in which a keyword appears on the first and second rows the last five letters of the normal alphabet, VWXYZ, will rarely be disturbed in sequence and will occupy the last row of the square, he "juggles" the letters given by the values tentatively established from frequency considerations, placing them in various positions in the square, together with VWXYZ, to correspond to the plain-text cipher relationships tentatively established. A later solution by Lieut. Frank Moorman, as described in Hitt's Manual, assumes that in a Playfair cipher prepared by means of a square in which the keyword occupies the first and second rows, if a digraphic frequency distribution is made, it will be found that the letters having the greatest combining power are very probably letters of the key. A still later solution, by Lieut. Commander Smith, is perhaps the most lucid and systematized of the three. He sets forth in definite language certain considerations which the other two writers certainly entertained but failed to indicate.

c. The following details have been summarized from Commander Smith's solution:

(1) The Playfair cipher may be recognized by virtue of the fact that it always contains an even number of letters, and that when divided into groups of two letters each, no group contains a repetition of the same letter, as NN or EE. Repetitions of digraphs, trigraphs, and polygraphs will be evident in fairly long messages.

(2) Using the square[8] shown in Fig. 25*a*, there are two general cases to be considered, as regards the results of encipherment:

B	A	N	K	R
D	E	F	G	H
I–J	L	M	O	Q
U	P	T	C	Y
S	V	W	X	Z

FIGURE 25*a*.

---

[6] This cipher was really invented by Sir Charles Wheatstone but receives its name from Lord Playfair, who apparently was its sponsor before the British Foreign Office. See Wemyss Reid, *Memoirs of Lyon Playfair*, London, 1899. A detailed description of this cipher will be found in Sec. VIII, *Advanced Military Cryptography*.

[7] Mauborgne, Lieut. J. O., U. S. A. *An advanced problem in cryptography and its solution*, Leavenworth, 1914.
Hitt, Captain Parker, U. S. A. *Manual for the solution of military ciphers*, Leavenworth, 1918.
Smith, Lieut. Commander W. W., U. S. N. In *Cryptography* by André Langie, translated by J. C. H. Macbeth, New York, 1922.

[8] The Playfair square accompanying Commander Smith's solution is based upon the keyword BANKRUPTCY, "to be distributed between the first and fourth lines of the square." This is a simple departure from the original Playfair scheme in which the letters of the keyword are written from left to right and in consecutive lines from the top downward.

CASE 1. Letters at opposite corners of a rectangle. The following illustrative relationships are found:

$$\begin{gathered} TH_p=YF_e \\ HT_p=FY_e \\ YF_p=TH_e \\ FY_p=HT_e \end{gathered}$$

Reciprocity is complete.

CASE 2. Two letters in the same line or column. The following illustrative relationships are found:

$$\begin{gathered} AN_p=NK_e \\ NA_p=KN_e \end{gathered}$$

But NK_p does not $=AN_e$, nor does $KN_p=NA_e$.

Reciprocity is only partial.

(3) The foregoing gives rise to the following:

RULE I. (*a*) Regardless of the position of the letters in the square, if

$$\begin{gathered} 1.2=3.4, \text{ then} \\ 2.1=4.3 \end{gathered}$$

(*b*) If 1 and 2 form opposite corners of a rectangle, the following equations obtain:

$$\begin{gathered} 1.2=3.4 \\ 2.1=4.3 \\ 3.4=1.2 \\ 4.3=2.1 \end{gathered}$$

(4) A letter considered as occupying a position in a row can be combined with but four other letters in the same row; the same letter considered as occupying a position in a column can be combined with but four other letters in the same column. Thus, this letter can be combined with only 8 other letters all told, under Case 2, above. But the same letter considered as occupying a corner of a rectangle can be combined with 16 other letters, under Case 1, above. Commander Smith derives from these facts the conclusion that "it would appear that Case 1 is twice as probable as Case 2." He continues thus (notation my own):

"Now in the square, note that:

$$\begin{array}{lcl} AN_p=NK_e & & EN_p=FA_e \\ GN_p=FK_e & & EM_p=FL_e \\ ON_p=MK_e & \text{also} & ET_p=FP_e \\ CN_p=TK_e & & EW_p=FV_e \\ XN_p=WK_e & & EF_p=FG_e \end{array}$$

"From this it is seen that of the 24 equations that can be formed when each letter of the square is employed either as the initial or final letter of the group, five will indicate a repetition of a corresponding letter of plain text.

"Hence, RULE II. After it has been determined, in the equation 1.2=3.4, that, say, $EN_p=FA_e$, there is a probability of one in five that any other group beginning with F_e indicates $E\Theta_p$, and that any group ending in A_e indicates ΘN_p.

"After such combinations as ER_p, OR_p, and EN_p have been assumed or determined, the above rule may be of use in discovering additional digraphs and partial words." [9]

Rule III. In the equation 1.2=3.4, 1 and 3 can never be identical, nor can 2 and 4 ever be identical. Thus, AN_p could not possibly be represented by AY_c, nor could ER_p be represented by KR_c. This rule is useful in elimination of certain possibilities when a specific message is being studied.

Rule IV. In the equation $1.2_p=3.4_c$, if 2 and 3 are identical, the letters are all in the same row or column, and in the relative order 124. In the square shown, $AN_p=NK_c$ and the absolute order is ANK. The relative order 124 includes five absolute orders which are cyclic permutations of one another. Thus: ANK.., NK..A, K..AN, ..ANK, and .ANK..

Rule V. In the equation $1.2_p=3.4_c$, if 1 and 4 are identical, the letters are all in the same row or column, and in the relative order 243. In the square shown, $KN_p=RK_c$ and the absolute order is NKR. The relative order 243 includes five absolute orders which are cyclic permutations of one another. Thus NKR.., KR..N, R..NK, ..NKR, and .NKR..

Rule VI. "Analyze the message for group recurrences. Select the groups of greatest recurrence and assume them to be high-frequency digraphs. Substitute the assumed digraphs throughout the message, testing the assumptions in their relation to other groups of the cipher. The reconstruction of the square proceeds simultaneously with the solution of the message and aids in hastening the translation of the cipher."

---

[9] There is an error in this reasoning. Take, for example, the 24 equations having F as an initial letter:

Case		Case		Case		Case	
1.	$FB_c=DN_p$	2.	FE=ED	2.	FT=NM	1.	FX=GW
2.	FD =EH	1.	FL=EM	2.	FW=NT	1.	FR=HN
1.	FI =DM	1.	FP=ET	1.	FK=GN	2.	FH=EG
1.	FU =DT	1.	FV=EW	2.	FG=EF	1.	FQ=HM
1.	FS =DW	2.	FN=NW	1.	FO=GM	1.	FY=HT
1.	FA =EN	2.	FM=NF	1.	FC=GT	1.	FZ=HW

Here, the initial letter F_c represents the following initial letters of plain-text digraphs:

$D\theta_p$, $E\theta_p$, $N\theta_p$, $G\theta_p$, and $H\theta_p$.

It is seen that F_c represents D_p, N_p, G_p, H_p 4 times each, and E_p, 8 times. Consequently, supposing that it has been determined that $FA_c=EN_p$, the probability that F_c will represent E_p is not 1 in 5 but 8 in 24, or 1 in 3; but supposing that it has been determined that $FW_c=NT_p$, the probability that F_c will represent N_p is 4 in 24 or 1 in 6. The difference in these probabilities is occasioned by the fact that the first instance, $FA_c=EN_p$ corresponds to a Case 1 encipherment, the second instance, $FW_c=NT_p$, to a Case 2 encipherment. But there is no way of knowing initially, and without other data, whether one is dealing with a Case 1 or Case 2 encipherment. Only as an approximation, therefore, may one say that the probability of F_c representing a given θ_p is 1 in 5. A probability of 1 in 5 is of almost trivial importance in this situation, since it represents such a "long shot" for success. The following rule might be preferable: If the equation 1.2=3.4 has been established, where all the letters represented by 1, 2, 3, and 4 are different, then there is a probability of 4/5 that a Case 1 encipherment is involved. Consequently, if at the same time another equation, 3.6=5.2, has been established, where 2 and 3 represent the same letters as in the first equation, and 5 and 6 are different letters, also different from 2 and 3, there is a probability of 16/25 that the equation 1.6=5.4 is valid; or if at the same time that the equation 1.2=3.4 has been determined, the equation 1.6=5.4 has also been established, then there is a probability of 16/25 that the equation 3.6=5.2 is valid. (Check this by noting the following equations based upon Fig. 25*a*: $\overset{1}{C}\overset{2}{E}=\overset{3}{P}\overset{4}{G}$, $\overset{3}{P}\overset{6}{H}=\overset{5}{Y}\overset{2}{E}$, $\overset{1}{C}\overset{6}{H}=\overset{5}{Y}\overset{4}{G}$. Note the positions occupied in Fig. 25*a* by the letters involved.) Likewise, if the equations 1.2=3.4 and 1.6=3.5 have been simultaneously established, then there is a probability that the equation 2.5=4.6 is valid; or if the equations 1.2=3.4 and 2.5=4.6 have been simultaneously established, then there is a probability that the equation 2.5=4.6 is valid. (Check this by noting the following equations: $\overset{1}{C}\overset{2}{E}=\overset{3}{P}\overset{4}{G}$; $\overset{1}{C}\overset{6}{A}=\overset{3}{P}\overset{5}{K}$; $\overset{2}{E}\overset{5}{K}=\overset{4}{G}\overset{6}{A}$; note the positions occupied in Fig. 25*a* by the letters involved.) However, it must be added that these probabilities are based upon assumptions which fail to take into account any considerations whatever as to frequency of letters or specificity of composition of the matrix. For instance, suppose the 5 high-frequency letters E, T, R, I, N all happen to fall in the same row or column in the matrix; the number of Case 2 encipherments would be much greater than expectancy and the probability that the equation 1.2=3.4 represents a Case 1 encipherment falls much below 4/5.

d. (1) When solutions for the Playfair cipher system were first developed, based upon the fact that the letters were inserted in the cells in keyword-mixed order, cryptographers thought it desirable to place stumbling blocks in the path of such solution by departing from strict, keyword-mixed order. Playfair squares of the latter type are designed as "modified Playfair squares." One of the simplest methods is illustrated in Fig. 25*a*, wherein it will be noted that the last five letters of the keyword proper are inserted in the fourth row of the square instead of the second, where they would naturally fall. Another method is to insert the letters within the cells from left to right and top downward but use a sequence that is a keyword-mixed sequence developed by a columnar transposition based upon the keyword proper. Thus, using the keyword BANKRUPTCY:

2	1	5	4	7	9	6	8	3	10
B	A	N	K	R	U	P	T	C	Y
D	E	F	G	H	I	L	M	O	Q
S	V	W	X	Z					

Sequence: A E V B D S C O K G X N F W P L R H Z T M U I Y Q

The Playfair square is as follows:

A	E	V	B	D
S	C	O	K	G
X	N	F	W	P
L	R	H	Z	T
M	U	I	Y	Q

FIGURE 25*b*.

(2) In the foregoing matrix practically all indications that the square has been developed from a keyword have disappeared. The principal disadvantage of such an arrangement is that it requires more time to locate the letters desired, both in cryptographing and decryptographing, than it usually does when a semblance of normal alphabetic order is preserved in the matrix.

(3) Note the following three squares:

Z	T	L	R	H
Y	Q	M	U	I
B	D	A	E	V
K	G	S	C	O
W	P	X	N	F

FIGURE 25*c*.

O	K	G	S	C
F	W	P	X	N
H	Z	T	L	R
I	Y	Q	M	U
V	B	D	A	E

FIGURE 25*d*.

N	F	W	P	X
R	H	Z	T	L
U	I	Y	Q	M
E	V	B	D	A
C	O	K	G	S

FIGURE 25*e*.

At first glance they all appear to be different, but closer examination shows them to be regular commutations or *cyclic permutations* of one another and of the square in Fig. 25*b*. They yield identical equivalents in all cases. However, if an attempt be made to reconstruct the original keyword, it would be much easier to do so from Fig. 25*b* than from any of the others, because in Fig. 25*b* the keyword-mixed sequence has not been disturbed as much as in Figs. *25c*, *d*, *e*. In working with Playfair ciphers, the student should be on the lookout for such instances of cyclic permutation of the original Playfair square, for during the course of solution he will not know whether he is building up the original or an equivalent cyclic permutation of the original matrix; only after he has completely reconstructed the matrix will he be able to determine this point.

(4) If the columns of a given Playfair square or matrix, M_1, are permuted cyclically (shifting column 5 to the left so as to change the order 12345 into 51234, then shifting column 4 to produce the order 45123, and so on), a set of five squares, M_1, M_2, M_3 M_4, and M_5, all having superficially different configurations but yielding identical equivalents will be obtained. From each of these five matrices, four more may be obtained by a homologous cyclic permutation applied to the rows. Thus a set of 25 superficially different but cryptographically equivalent matrices may be produced. By interchanging the rows and the columns of the original square, M_1 (i. e., making column 1 into row 1, column 2 into row 2, and so on), a new matrix, M'_1, is produced, in which all Case 2 encipherments (letters in the same row or in the same column) will be identical with those of the original M_1 matrix, but all Case 1 encipherments (letters at diagonally opposite corners of a rectangle) will be the reverse of those of the original M_1 square. (Note the effect of turning a Playfair square 90° or 270°.) By cyclically permuting the rows and then the columns of the M'_1, matrix, another set of 25 superficially different but cryptographically equivalent matrices may be produced. Thus, 49 matrices are derivable by cyclic permutation of an original square, making a total of 50 matrices in all. Now, the Case 2 encipherments obtained from all 50 of these matrices will be identical but Case 1 encipherments by means of any one of the 25 matrices of the first set will be the reverse of those obtained by means of any one of the 25 matrices of the second set. Consequently, there are only 24 secondary matrices which may be derived by cyclic permutation from a basic or primary Playfair square, and which are cryptographically equivalent among themselves and with the original square, thus making, in all, 25 cryptographically equivalent matrices. The usefulness of this property of cryptographic equivalence inherent in cyclically related matrices, as well as the property of reversibility as regards Case 1 encipherments from two matrices bearing a 90° or 270° phase relationship, will become clear in the subsequent paragraphs, when the mechanics of the reconstruction of a Playfair square are presented.

e. (1) The steps in the solution of a typical example of this cipher may be useful. Let the message be as follows:

```
    1 2 3 4 5   6 7 8 9 10  11 12 13 14 15  16 17 18 19 20  21 22 23 24 25  26 27 28 29 30
A.  V T Q E U   H I O F T   C H X S C   A K T V T   R A Z E V   T A G A E
B.  O X T Y M   H C R L Z   Z T Q T D   U M C Y C   X C T G M   T Y C Z U
C.  S N O P D   G X V X S   C A K T V   T P K P U   T Z P T W   Z F N B G
D.  P T R K X   I X B P R   Z O E P U   T O L Z E   K T T C S   N H C Q M
E.  V T R K M   W C F Z U   B H T V Y   A B G I P   R Z K P C   Q F N L V
F.  O X O T U   Z F A C X   X C P Z X   H C Y N O   T Y O L G   X X I I H
G.  T M S M X   C P T O T   C X O T T   C Y A T E   X H F A C   X X C P Z
H.  X H Y C T   X W L Z T   S G P Z T   V Y W C E   T W G C C   M B H M Q
J.  Y X Z P W   G R T I V   U X P U M   Q R K M W   C X T M R   S W G H B
K.  X C P T O   T C X O T   M I P Y D   N F G K I   T C O L X   U E T P X
L.  X F S R S   U Z T D B   H O Z I G   X R K I X   Z P P V Z   I D U H Q
M.  O T K T K   C C H X X
```

(2) Without going through the preliminary tests in detail, with which it will be assumed that the student is now familiar,[10] the conclusion is reached that the cryptogram is digraphic in nature, and a digraphic frequency distribution is made (Fig. 26).

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A					/		/																		
B							//	//																	
C	//				/	/		//				/				/	/						𝍸	/	
D		/					/						/							//					
E															/				/						
F	//						/						//						/						
G			/																				//		
H		/	/											/		/									
I								/						/	/						/		/		
K			/						/						/				////						
L																					/				/
M			/					/	/							//			/			//			
N														/											
O											///				/				𝍸 /				//		
P										/							/		////	//	/		/	/	///
Q					/							/							/						
R	/									////								/	/						/
S							/					/	//				/			/					
T			///		/		/					//									//	/	/	//	/
U								/											/				/		/
V																			𝍸 /						
W							//				/														/
X		/	𝍸 /			/		///	//									//		/	/				
Y	//		///																			/	/		
Z					//				//					/	//				///	//					

FIGURE 26.

---

[10] See Par. 44c.

Since there are no double-letter groups, the conclusion is reached that a Playfair cipher is involved and the message is rewritten in digraphs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	VT	QE	UH	IO	FT	CH	XS	CA	KT	VT	RA	ZE	VT	AG	AE
B.	OX	TY	MH	CR	LZ	ZT	QT	DU	MC	YC	XC	TG	MT	YC	ZU
C.	SN	OP	DG	XV	XS	CA	KT	VT	PK	PU	TZ	PT	WZ	FN	BG
D.	PT	RK	XI	XB	PR	ZO	EP	UT	OL	ZE	KT	TC	SN	HC	QM
E.	VT	RK	MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP	CQ	FN	LV
F.	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY	OL	GX	XI	IH
G.	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ
H.	XH	YC	TX	WL	ZT	SG	PZ	TV	YW	CE	TW	GC	CM	BH	MQ
J.	YX	ZP	WG	RT	IV	UX	PU	MQ	RK	MW	CX	TM	RS	WG	HB
K.	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET	PX
L.	XF	SR	SU	ZT	DB	HO	ZI	GX	RK	IX	ZP	PV	ZI	DU	HQ
M.	OT	KT	KC	CH	XX										

(3) The following three fairly lengthy repetitions are noted:

Lines									
F.	OT	UZ	FA	CX	XC	PZ	XH	CY	NO
G.	TE	XH	FA	CX	XC	PZ	XH	YC	TX
A.	FT	CH	XS	CA	KT	VT	RA	ZE	
C.	DG	XV	XS	CA	KT	VT	PK	PU	
G.	TM	SM	XC	PT	OT	CX	OT	TC	
K.	WG	HB	XC	PT	OT	CX	OT	MI	

The first long repetition, with the sequent reversed digraphs CX and XC immediately suggests the word BATTALION, split up into –B AT TA LI ON and the sequence containing this repetition in lines F and G becomes as follows:

Line F....................	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
				B	AT	TA	LI	ON			
Line G....................	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
			ON	B	AT	TA	LI	ON			

(4) Because of the frequent use of numerals before the word BATTALION and because of the appearance of ON before this word in line G, the possibility suggests itself that the word before BATTALION in line G is either ONE or SECOND. The identical digraph FA in both cases gives a hint that the word BATTALION in line F may also be preceded by a numeral; if ONE is

correct in line G, then THREE is possible in line F. On the other hand, if SECOND is correct in line G, then THIRD is possible in line F. Thus:

Line F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
1st hypothesis	—	TH	RE	EB	AT	TA	LI	ON			
2nd hypothesis	—	TH	IR	DB	AT	TA	LI	ON			
Line G	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
1st hypothesis	—	—	ON	EB	AT	TA	LI	ON			
2nd hypothesis	-S	EC	ON	DB	AT	TA	LI	ON			

First, note that if either hypothesis is true, then $OT_c=TH_p$. The frequency distribution shows that OT occurs 6 times and is in fact the most frequent digraph in the message. Moreover, by Rule I of subparagraph *b*, if $OT_c=TH_p$ then $TO_c=HT_p$. Since HT_p is a very rare digraph in normal plain text, TO_c should either not occur at all in so short a message or else it should be very infrequent. The frequency distribution shows its entire absence. Hence, there is nothing inconsistent with the possibility that the word in front of BATTALION in line F is THREE or THIRD, and some evidence that it is actually one or the other.

(5) But can evidence be found for the support of one hypothesis against the other? Let the frequency distribution be examined with a view to throwing light upon this point. If the first hypothesis is true, then $UZ_c=RE_p$, and, by Rule I, $ZU_c=ER_p$. The frequency distribution shows but one occurrence of UZ_c and but two occurrences of ZU_c. These do not look very good for RE and ER. On the other hand, if the second hypothesis is true, then $UZ_c=IR_p$ and, by Rule I, $ZU_c=RI_p$. The frequencies are much more favorable in this case. Is there anything inconsistent with the assumption, on the basis of the second hypothesis, that $TE_c=EC_p$? The frequency distribution shows no inconsistency, for TE_c occurs once and ET_c ($=CE_p$, by Rule I) occurs once. As regards whether $FA_c=EB_p$ or DB_p, both hypotheses are tenable; possibly the second hypothesis is a shade better than the first, on the following reasoning: By Rule I, if $FA_c=EB_p$ then $AF_c=BE_p$, or if $FA_c=DB_p$ then $AF_c=BD_p$. The fact that no AF_c occurs, whereas at least one BE_p may be expected in this message, inclines one to the second hypothesis, since BD_p is very rare.

(6) Let the 2nd hypothesis be assumed to be correct. The additional values are tentatively inserted in the text, and in lines G and K two interesting repetitions are noted:

Line G	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ	XH
			TA		TH	AT	TH		-S	EC	ON	DB	AT	TA	LI	ON
Line K	WG	HB	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET
			TA		TH	AT	TH									

This certainly looks like STATE THAT THE. . ., which would make $TE_p=PT_c$. Furthermore, in line G the sequence STATETHATTHE. . SECONDBATTALION can hardly be anything else than STATE THAT THEIR SECOND BATTALION, which would make $TC_c=EI_p$ and $YA_c=RS_p$. Also $SM_c=-S_p$.

(7) It is perhaps high time that the whole list of tentative equivalent values be studied in relation to their consistency with the positions of letters in the Playfair square; moreover, by so doing, additional values may be obtained in the process. The complete list of values is as follows:

Assumed values	*Derived by Rule I*
$AT_p=CX_c$	$TA_p=XC_c$
$LI_p=PZ_c$	$IL_p=ZP_c$
$ON_p=XH_c$	$NO_p=HX_c$
$TH_p=OT_c$	$HT_p=TO_c$
$IR_p=UZ_c$	$RI_p=ZU_c$
$DB_p=FA_c$	$BD_p=AF_c$
$EC_p=TE_c$	$CE_p=ET_c$
$TE_p=PT_c$	$ET_p=TP_c$
$EI_p=TC_c$	$IE_p=CT_c$
$RS_p=YA_c$	$SR_p=AY_c$
$-S_p=SM_c$	$S-_p=MS_c$

(8) By Rule V, the equation $TH_p=OT_c$ means that H, T, and O are all in the same row or column and in the relative order 2–4–3; similarly, C, E, and T are in the same row or column and in the relative order 243. Further E, P, and T are in the same row and column, and their relative order is also 243. That is, these sequences must occur in the square:

(1)	(2)	(3)
H T O . . , or	C E T . . , or	E T P . . , or
T O . . H , or	E T . . C , or	T P . . E , or
O . . H T , or	T . . C E , or	P . . E T , or
. . H T O , or	. . C E T , or	. . E T P , or
. H T O .	. C E T .	. E T P .

(9) Noting the common letters E and T in the second and third sets of relative orders, these may be combined into one sequence of four letters. Only one position remains to be filled and noting, in the list of equivalents that $EI_p=TC_c$, it is obvious that the letter I belongs to the CET sequence. The complete sequence is therefore as follows:

C E T P I , or
E T P I C , or
T P I C E , or
P I C E T , or
I C E T P

(10) Taking up the HTO sequence, it is noted, in the list of equivalents that $ON_p=XH_c$, an equation containing two of the three letters of the HTO sequence. From this it follows that N and X must belong to the same row or column as HTO. The arrangement must be one of the following:

H T O X N
T O X N H
O X N H T
X N H T O
N H T O X

(11) Since the sequence containing HTOXN has a common letter (T) with the sequence CETPI, it follows that if the HTOXN sequence occupies a row, then the CETPI sequence must occupy a column; or, if the HTO sequence occupies a column, then the CETPI sequence must

occupy a row; and they may be combined by means of their common letter, T. According to subpar. *d* (4), the two sequences may be inserted within a Playfair square in 25 different ways by cyclically permuting and shifting the letters of one of these two sequences; and the same two sequences may be again inserted in another set of 25 ways by cyclically permuting and shifting the letters of the other of these two sequences. In Fig. 27 the diagrams labeled (1) to (10), inclusive, show 10 of the possible 25 obtainable by making the HTOXN sequence one of the rows of the square; diagrams (11) and (12) show 2 of the possible 25 obtainable by making the HTOXN sequence one of the columns of the square. The entire complement of 25 arrangements for each set may easily be drawn up by the student; space forbids their being completely set forth and it is really unnecessary to do so.

(1)

C				
E				
T	O	X	N	H
P				
I				

(2)

	C			
	E			
H	T	O	X	N
	P			
	I			

(3)

		C		
		E		
N	H	T	O	X
		P		
		I		

(4)

			C	
			E	
X	N	H	T	O
			P	
			I	

(5)

				C
				E
O	X	N	H	T
				P
				I

(6)

P				
I				
C				
E				
T	O	X	N	H

(7)

	P			
	I			
	C			
	E			
H	T	O	X	N

(8)

		P		
		I		
		C		
		E		
N	H	T	O	X

(9)

			P	
			I	
			C	
			E	
X	N	H	T	O

(10)

				P
				I
				C
				E
O	X	N	H	T

(11)

				N
				H
P	I	C	E	T
				O
				X

(12)

N				
H				
T	P	I	C	E
O				
X				

FIGURE 27.

(12) Before trying to discover means whereby the actual or primary matrix may be detected from among the full set of 50 possible matrices, the question may be raised: is it necessary? So far as concerns Case 2 encipherments, since any one of the 50 arrangements will yield the same equivalents as any of the remaining 49, perhaps a relative arrangement will do. The configurations shown in Fig. 27 constitute what may be termed *skeleton matrices*, and one of them will be selected for experiment.

(13) Let skeleton matrix 8 be arbitrarily selected for trial.

		P		
		I		
		C		
		E		
N	H	T	0	X

FIGURE 28a.

(14) What additional letters can be inserted, using as a guide the list of equivalents in subparagraph (7)? There is $AT_p=CX_c$, for example. It contains only one letter, A, not in the skeleton matrix selected for trial, and this letter may immediately be placed, as shown:[10]

		P		
		I		
		C		A
		E		
N	H	T	0	X

FIGURE 28b.

Scanning the list for additional cases of this type, none are found. But seeing that several high-frequency letters have already been inserted in the matrix, perhaps reference to the cryptogram itself in connection with values derived from these inserted letters may yield further clues. For example, the vowels A, E, I, and 0 are all in position, as are the very frequent consonants N and T. The following combinations may be studied:

$AN_p=\theta X_c$ $AT_p=CX_c$ $NA_p=X\theta_c$ $TA_p=XC_c$
$EN_p=\theta T_c$ $ET_p=TP_c$ $NE_p=T\theta_c$ $TE_p=PT_c$
$IN_p=\theta T_c$ $IT_p=CP_c$ $NI_p=T\theta_c$ $TI_p=PC_c$
$ON_p=XH_c$ $OT_p=XO_c$ $NO_p=HX_c$ $TO_p=OX_c$

$AT_p(=CX_c)$, $TA_p(=XC_c)$, $ON_p(=XH_c)$, $TE_p(=PT_c)$ and $ET_p(=TP_c)$ have already been inserted in the text. Of the others, only $OX_c(=TO_p)$ occurs two times, and this value can be at once inserted in the text. But can the equivalents of AN, EN, or IN be found from frequency considerations?

---

[10] The fact that the placement of A yields $AT_p=CX_c$ means that the skeleton matrix selected for experiment really belongs to the correct set of 25 possible cylic permutations, and that the letters of the NHTOX sequence belong in a row, the letters of the PICET sequence belong in a column of the original Playfair square. If the reverse were the case, one could not obtain $AT_p=CX_c$ but would obtain $AT_p=XC_c$. Thus the equation $AT_p=XC_c$ unequivocally distinguishes the skeleton matrices 1 to 10, inclusive (Fig. 27), from skeleton matrices 11 and 12, since if either of the latter had been selected instead of skeleton matrix 8, the letter A could not have been positioned to satisfy this equation without contradiction.

Take EN_p, for example; it is represented by θT_c. What combination of θT is most likely to represent EN_p among the following candidates:

KT_c (4 times); by Rule I, NE_p would$=TK_c$ (no occurrences)

VT_c (5 times); by Rule I, NE_p would$=TV_c$ (2 times)

ZT_c (3 times); by Rule I, NE_p would$=TZ_c$ (1 time)

VT_c certainly looks good: it begins the message, suggesting the word ENEMY; in line H, in the sequence PZTV would become LINE. Let this be assumed to be correct, and let the word ENEMY also be assumed to be correct. Then $EM_p=QE_c$ and the square then becomes as shown herewith:

		P		
		I		
		C		A
V	M	E	Q	
N	H	T	O	X

FIGURE 28c.

(15) In line E is seen the following sequence:

Line E: ....VT	RK	MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP	CQ	FN	LV
EN				RI		NE	RS		PT			E		

The sequence ...RI..NERS..PT... suggests PRISONERS CAPTURED, as follows:

MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP
	P	RI	SO	NE	RS	CA	PT	UR	ED

This gives the following new values: $\theta P_p=CF_c$; $SO_p=BH_c$; $CA_p=BG_c$; $UR_p=RZ_c$; $ED_p=KP_c$.

The letters B and G can be placed in position at once, since the positions of C and A are already known. The insertion of the letter B immediately permits the placement of the letter S, from the equation $SO_p=BH_c$. Of the remaining equations only $ED_p=KP_c$ can be used. Since E and P are fixed and are in the same column, D and K must be in the same column, and moreover the K must be in the same row as E. There is only one possible position for K, viz, immediately after Q. This automatically fixes the position of D. The square is now as shown herewith:

		P		D
		I		
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

FIGURE 28d.

(16) A review of all equations, including the very first ones established, gives the following which may now be used: $DB_p=FA_c$; $RS_p=YA_c$. The first permits the immediate placement of F; the second, by elimination of possible positions, permits the placement of both R and Y. The square is now as shown herewith:

		P	F	D
	Y	I		R
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

FIGURE 28e.

Once more a review is made of all remaining thus far unused equations. $LI_p=PZ_c$ now permits the placement of L and Z. $IR_p=UZ_c$ now permits the placement of U, which is confirmed by the equation $UR_p=RZ_c$ from the word CAPTURED.

L		P	F	D
Z	Y	I	U	R
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

FIGURE 28f.

There is then only one cell vacant, and it must be occupied by the only letter left unplaced, viz, W. Thus the whole square has been reconstructed, and the message can now be decryptographed.

(17) Is the square just reconstructed identical with the original, or is it a cyclic permutation of a keyword-mixed Playfair square of the type illustrated in Fig. 25b? Even though the message can be read with ease, this point is still of interest. Let the sequence be written in five ways, each composed of five partial sequences made by cyclicly permuting each of the horizontal rows of the reconstructed square. Thus:

	Row 1	Row 2	Row 3	Row 4	Row 5
(a)	L W P F D	Z Y I U R	G S C B A	V M E Q K	N H T O X
(b)	W P F D L	Y I U R Z	S C B A G	M E Q K V	H T O X N
(c)	P F D L W	I U R Z Y	C B A G S	E Q K V M	T O X N H
(d)	F D L W P	U R Z Y I	B A G S C	Q K V M E	O X N H T
(e)	D L W P F	R Z Y I U	A G S C B	K V M E Q	X N H T O

By experimenting with these five sequences, in an endeavor to reconstruct a transposition rectangle conformable to a keyword sequence, the last sequence yields the following:

	P	Y	A	C	M	N
D	F	I	G	B	E	H
L	R	U	S	K	Q	T
W	Z			V	X	O

By shifting the O from the last position to the first, and rearranging the columns, the following is obtained:

2	5	3	6	1	4	7
C	O	M	P	A	N	Y
B	D	E	F	G	H	I
K	L	Q	R	S	T	U
V	W	X	Z			

The original square must have been this:

A	G	S	C	B
K	V	M	E	Q
X	N	H	T	O
D	L	W	P	F
R	Z	Y	I	U

FIGURE 28g

f. Continued practice in the solution of Playfair ciphers will make the student quite expert in the matter and will enable him to solve shorter and shorter messages.[11] Also, with practice it will become a matter of indifference to him as to whether the letters are inserted in the square with any sort of regularity, such as simple keyword-mixed order, columnar transposed keyword-mixed order, or in a purely random order.

g. It may perhaps seem to the student that the foregoing steps are somewhat too artificial, a bit too "cut and dried" in their accuracy to portray the process of analysis, as it is applied in practice. For example, the critical student may well object to some of the assumptions and the reasoning in step (5) above, in which the words THREE and ONE (1st hypothesis) were rejected in favor of the words THIRD and SECOND (2nd hypothesis). This rested largely upon the rejection of RE$_p$ and ER$_p$ as the equivalents of UZ$_c$ and ZU$_c$, and the adoption of IR$_p$ and RI$_p$ as their equivalents. Indeed, if the student will examine the final message with a critical eye he will find that while the bit of reasoning in step (5) is perfectly logical, the assumption upon which it is based is in fact wrong, for it happens that in this case ER$_p$ occurs only once and RE$_p$ does not occur at all. Consequently, although most of the reasoning which led to the rejection of the 1st hypothesis and the adoption of the 2nd was logical, it was in fact based upon erroneous assumption. In other words, despite the fact that the assumption was incorrect, a correct deduction was made. *The student should take note that in cryptanalysis situations of this sort are not at all unusual.* Indeed they are to be expected and a few words of explanation at this point may be useful.

---

[11] The author once had a student who "specialized" in Playfair ciphers and became so adept that he could solve messages containing as few as 50–60 letters within 30 minutes.

h. Cryptanalysis is a science in which deduction, based upon observational data, plays a very large role. But it is also true that in this science most of the deductions usually rest upon assumptions. It is most often the case that the cryptanalyst is forced to make his assumptions upon a quite limited amount of text. It cannot be expected that assumptions based upon statistical generalizations will always hold true when applied to data comparatively very much smaller in quantity than the total data used to derive the generalized rules. Consequently, as regards assumptions made in specific messages, *most of the time* they will be correct, but *occasionally* they will be incorrect. In cryptanalysis it is often found that among the correct deductions there will be cases in which subsequently discovered facts do not bear out the assumptions on which the deduction was based. Indeed, it is sometimes true that if the *facts* had been known *before* the deduction was made, this knowledge would have prevented making the correct deduction. For example, suppose the cryptanalyst had somehow or other divined that the message under consideration contained no RE, only one ER, one IR, and two RI's (as is actually the case). He would certainly not have been able to choose between the words THREE and ONE (1st hypothesis) as against THIRD and SECOND (2d hypothesis). But because he assumes that there should be more ER_p's and RE_p's than IR's and RI's in the message, he deduces that UZ_c cannot be RE_p, rejects the 1st hypothesis and takes the 2d. It later turns out, after the problem has been solved, that the deduction was correct, although the assumption on which it was based (expectation of more frequent appearance of RE_p and ER_p) was, in fact, *not* true in this particular case. The cryptanalyst can only hope that the number of times when his deductions are correct, even though based upon assumptions which later turn out to be erroneous, will abundantly exceed the number of times when his deductions are wrong, even though based upon assumptions which later prove to be correct. If he is lucky, the making of an assumption which is really not true will make no difference in the end and will not delay solution; but if he is specially favored with luck, it may actually help him solve the message—as was the case in this particular example.

i. Another comment of a general nature may be made in connection with this specific example. The student may ask what would have been the procedure in this case if the message had not contained such a tell-tale repetition as the word BATTALION, which formed the point of departure for the solution, or, as it is often said, permitted an "entering wedge" to be driven into the message. The answer to his query is that if the word BATTALION had not been repeated, there would probably have been some other repetition which would have permitted the same sort of attack. If the student is looking for cut and dried, straight-forward, unvarying methods of attack, he should remember that cryptanalysis, while it may be considered a branch of mathematics, is not a science which has many "general solutions" such as are found and expected in mathematics proper. It is inherent in the very nature of cryptanalytics that, *as a rule*, only general principles can be established; their practical application must take advantage of peculiarities and particular situations which are noted in specific messages. This is especially true in a text on the subject. The illustration of a general principle requires a specific example, and the latter must of necessity manifest characteristics which make it different from any other example. The word BATTALION was not purposely repeated in this example in order to make the demonstration of solution easy; "it just happened that way." In another example, some other entering wedge would have been found. The student can be expected to learn only the *general principles* which will enable him to take advantage of the *specific characteristics* manifested in *specific cases.* Here it is desired to illustrate the general principles of solving Playfair ciphers and to point out the fact that entering wedges must and can be found. The specific nature of the entering wedge varies with specific examples.

SECTION X

CONCLUDING REMARKS

Paragraph

Special remarks concerning the initial classification of cryptograms........ 47

Ciphers employing characters other than letters or figures........ 48

Concluding remarks concerning monoalphabetic substitution........ 49

Analytical key for cryptanalysis........ 50

47. Special remarks concerning the initial classification of cryptograms.—*a*. The student should by this time have a good conception of the basic nature of monoalphabetic substitution and of the many "changes" which may be rung upon this simple tune. The first step of all, naturally, is to be able to classify a cryptogram properly and place it in either the transposition or the substitution class. The tests for this classification have been given and as a rule the student will encounter no difficulty in this respect.

b. There are, however, certain kinds of cryptograms whose class cannot be determined in the usual manner, as outlined in Par. 13 of this text. First of all there is the type of code message which employs bona-fide dictionary words as code groups.[1] Naturally, a frequency distribution of such a message will approximate that for normal plain text. The appearance of the message, however, gives clear indications of what is involved. The study of such cases will be taken up in its proper place. At the moment it is only necessary to point out that these are *code* messages and not *cipher*, and it is for this reason that in Pars. 12 and 13 the words "cipher" and "cipher messages" are used, the word "cryptogram" being used only where technically correct.

c. Secondly, there come the unusual and borderline cases, including cryptograms whose nature and type can *not* be ascertained from frequency distributions. Here, the cryptograms are technically not ciphers but special forms of disguised secret writings which are rarely susceptible of being classed as transposition or substitution. These include a large share of the cases wherein the cryptographic messages are disguised and carried under an external, innocuous text which is innocent and seemingly without cryptographic content—for instance, in a message wherein specific letters are indicated in a way not open to suspicion under censorship, these letters being intended to constitute the letters of the cryptographic message and the other letters constituting "dummies." Obviously, no amount of frequency tabulations will avail a competent, expert cryptanalyst in demonstrating or disclosing the presence of a cryptographic message, written and secreted within the "open" message, which serves but as an envelop and disguise for its authentic or real import. Certainly, such frequency tabulations can disclose the existence *neither* of substitution *nor* transposition in these cases, since both forms are absent. Another very popular method that resembles the method mentioned above has for its basis a simple grille. The whole words forming the secret text are inserted within perforations cut in the paper and the remaining space filled carefully, using "nulls" and "dummies", making a seemingly innocuous, ordinary message. There are other methods of this general type which can obviously neither be detected nor cryptanalyzed, using the principles of frequency of recurrences and repetition. These can not be further discussed herein, but at a subsequent date a special text may be written for their handling.[2]

---

[1] See Sec. XV, *Elementary Military Cryptography.*

[2] The subparagraph which the student has just read (47*c*) contains a hidden cryptographic message. With the hints given in Par. 35*e* let the student see if he can find it.

48. Ciphers employing characters other than letters or figures.—*a.* In view of the foregoing remarks, when so-called symbol ciphers, that is, ciphers employing peculiar symbols, signs of punctuation, diacritical marks, figures of "dancing men", and so on are encountered in practical work nowadays, they are almost certain to be simple, monoalphabetic ciphers. They are adequately described in romantic tales,[3] in popular books on cryptography, and in the more common types of magazine articles. No further space need be given ciphers of this type in this text, not only because of their simplicity but also because they are encountered in military cryptography only in sporadic instances, principally in censorship activities. Even in the latter cases, it is usually found that such ciphers are employed in "intimate" correspondence for the exchange of sentiments that appear less decorous when set forth in plain language. They are very seldom used by authentic enemy agents. When such a cipher is encountered nowadays it may practically always be regarded as the work of the veriest tyro, when it is not that of a "crank" or a mentally-deranged person.

b. The usual preliminary procedure in handling such cases, where the symbols may be somewhat confusing to the mind because of their unfamiliar appearance to the eye, is to substitute letters for them consistently throughout the message and then treat the resulting text as an ordinary cryptogram composed of letters is treated. This procedure also facilitates the construction of the necessary frequency distributions, which would be tedious to construct by using symbols.

c. A final word must be said on the subject of symbol ciphers by way of caution. When symbols are used to replace letters, syllables, and entire words, then the systems approach code methods in principle, and can become difficult of solution.[4] The logical extension of the use of symbols in such a form of writing is the employment of arbitrary characters for a specially developed "shorthand" system bearing little or no resemblance to well-known, and therefore nonsecret, systems of shorthand, such as Gregg, Pitman, etc. Unless a considerable amount of text is available for analysis, a privately-devised shorthand may be very difficult to solve. Fortunately, such systems are rarely encountered in military cryptography. They fall under the heading of cryptographic curiosities, of interest to the cryptanalyst in his leisure moments.[5]

d. In practical cryptography today, as has been stated above, the use of characters other than the 26 letters of the English alphabet is comparatively rare. It is true that there are a few governments which still adhere to systems yielding cryptograms in groups of figures. These are almost in every case code systems and will be treated in their proper place. In some cases cipher systems, or systems of enciphering code are used which are basically mathematical in character and operation, and therefore use numbers instead of letters. Some persons are inclined toward the use of numbers rather than letters because numbers lend themselves much more readily to certain arithmetical operations such as addition, subtraction, and so on, than do letters.[6] But there is usually added some final process whereby the figure groups are converted into letter groups, for the sake of economy in transmission.

---

[3] The most famous: Poe's *The Gold Bug;* Arthur Conan Doyle's *The Sign of Four.*

[4] The use of symbols for abbreviation and speed in writing goes back to the days of antiquity. Cicero is reported to have drawn up "a book like a dictionary, in which he placed before each word the notation (symbol) which should represent it, and so great was the number of notations and words that whatever could be written in Latin could be expressed in his notations."

[5] An example is found in the famous Pepys Diary, which was written in shorthand, purely for his own eyes by Samuel Pepys (1633–1703). "He wrote it in Shelton's system of tachygraphy (1641), which he complicated by using foreign languages or by varieties of his own invention whenever he had to record passages least fit to be seen by his servants, or by 'all the world.' "

[6] But, this of course, is because we are taught arithmetic by using numbers, based upon the decimal system as a rule. By special training one *could* learn to perform the usual "arithmetical" operations using letters. For example, using our English alphabet of 26 letters, where $A=1$, $B=2$, $C=3$, etc., it is obvious that $A+B=C$, just as $1+2=3$; $(A+B)^2=I$, etc. This sort of cryptographic arithmetic could be learned by rote, just as multiplication tables are learned.

e. The only notable exceptions to the statement contained in the first sentence of the preceding subparagraph are those of Russian messages transmitted in the Russian Morse alphabet and Japanese messages transmitted in the Kata Kana Morse alphabet. As regards Chinese, which is not an alphabetical language and comprises some 40,000 ideographs, since the Morse telegraph code comprises only some 40 combinations, telegrams in Chinese are usually prepared by means of codes which permit of substituting arbitrarily-assigned code groups for the characters. Usually the code groups consist of figures. One such code known as the *Official Chinese Telegraph Code*, has about 10,000 4-figure groups, beginning with 0001, and these are arranged so that there are 100 characters on each page. Sometimes, for purposes of secrecy or economy, these figure groups are enciphered and converted in letter groups.

49. Concluding remarks concerning monoalphabetic substitution.—*a.* The alert student will have by this time gathered that the solution of monoalphabetic substitution ciphers of the simple or fixed type are particularly easy to solve, once the underlying principles are thoroughly understood. As in other arts, continued practice with examples leads to facility and skill in solution, especially where the student concentrates his attention upon traffic all of the same general nature, so that the type of text which he is continually encountering becomes familiar to him and its peculiarities or characteristics of construction give clues for short cuts to solution. It is true that a knowledge of the general phraseology of messages, the kind of words used, their sequences, and so on, is of very great assistance in practical work in all fields of cryptanalysis. The student is urged to note particularly these finer details in the course of his study.

b. Another thing which the student should be on the lookout for in simple monoalphabetic substitution is the consecutive use of several different mixed cipher alphabets in a single long message. Obviously, a single, composite frequency distribution for the whole message will not show the characteristic crest and trough appearance of a simple monoalphabetic cipher, since a given cipher letter will represent different plain-text letters in different parts of the message. But if the cryptanalyst will carefully observe the distribution *as it is being compiled*, he will note that at first it presents the characteristic crest and trough appearance of monoalphabeticity, and that after a time it begins to lose this appearance. If possible he should be on the lookout for some peculiarity of grouping of letters which serves as an indicator for the shift from one cipher alphabet to the next. If he finds such an indicator he should begin a second distribution from that point on, and proceed until another shift or indicator is encountered. By thus isolating the different portions of the text, and restricting the frequency distributions to the separate monoalphabets, the problem may be treated then as an ordinary simple monoalphabetic substitution. Consideration of these remarks in connection with instances of this kind leads to the comment that it is often more advisable for the cryptanalyst to compile his own data, than to have the latter prepared by clerks, especially when studying a system *de novo*. For observations which will certainly escape an untrained clerk can be most useful and may indeed facilitate solution. For example, in the case under consideration, if a clerk should merely hand the uniliteral distribution to the cryptanalyst, the latter might be led astray; the appearance of the composite distribution might convince him that the cryptogram is a good deal more complicated than it really is.

c. Monoalphabetic substitution with variants represents an extension of the basic principle, with the intention of masking the characteristic frequencies resulting from a strict monoalphabeticity, by means of which solutions are rather readily obtained. Some of the subterfuges applied on the establishment of variant or multiple values are simple and more or less fail to serve the purpose for which they are intended; others, on the contrary, may interpose serious difficulties to a straightforward solution. But in no case may the problem be considered of more than ordinary difficulty. Furthermore, it should be recognized that where these subterfuges are really adequate to the purpose, the complications introduced are such that the practical manipulation of the system becomes as difficult for the cryptographer as for the cryptanalyst.

d. As already mentioned in monoalphabetic substitution with variants it is most common to employ figures or groups of figures. The reason for this is that the use of numerical groups seems more natural or easier to the uninitiated than does the use of varying combinations of letters. Moreover, it is easy to draw up cipher alphabets in which some of the letters are represented by single digits, others by pairs of digits. Thus, the decomposition of the cipher text which is an irregular intermixture of uniliteral and multiliteral equivalents, is made more complicated and correspondingly difficult for the cryptanalyst, who does not known which digits are to be used separately, which in pairs.

e. A few words may be added here in regard to a method which often suggests itself to laymen. This consists in using a book possessed by all the correspondents and indicating the letters of the message by means of numbers referring to specific letters in the book. One way consists in selecting a certain page and then giving the line number and position of the letter in the line, the page number being shown by a single initial indicator. Another way is to use the entire book, giving the cipher equivalents in groups of three numbers representing page, line, and number of letter. (Ex.: 75–8–10 means page 75, 8th line, 10th letter in the line.) Such systems are, however, extremely cumbersome to use and, when the cryptographing is done carelessly, can be solved. The basis for solution in such cases rests upon the use of adjacent letters on the same line, the accidental repetitions of certain letters, and the occurrence of unenciphered words in the messages, when laziness or fatigue intervenes in the cryptographing.[7]

f. It may also be indicated that human nature and the fallibility of cipher clerks is such that it is rather rare for an encipherer to make full use of the complement of variants placed at his disposal. The result is that in most cases certain of the equivalents will be used so much more often than others that diversities in frequencies will soon manifest themselves, affording important data for attack by the cryptanalyst.

g. In the World War the cases where monoalphabetic substitution ciphers were employed in actual operations on the Western Front were exceedingly rare because the majority of the belligerents had a fair knowledge of cryptography. On the Eastern Front, however, the extensive use, by the poorly prepared Russian Army, of monoalphabetic ciphers in the fall of 1914 was an important, if not the most important, factor in the success of the German operations during the Battle of Tannenberg.[8] It seems that a somewhat more secure cipher system was authorized, but proved too difficult for the untrained Russian cryptographic and radio personnel. Consequently, recourse was had to simple substitution ciphers, somewhat interspersed with plain text, and sometimes to messages completely in plain language. The damage which this faulty use of cryptography did to the Russian Army and thus to the Allied cause is incalculable.

h. Many of the messages found by censors in letters sent by mail during the World War were cases of monoalphabetic substitution, disguised in various ways.

---

[7] In 1915 the German Government conspired with a group of Hindu revolutionaries to stir up a rebellion in India, the purpose being to cause the withdrawal of British troops from the Western Front. Hindu conspirators in the United States were given money to purchase arms and ammunition and to transport them to India. For communication with their superiors in Berlin the conspirators used, among others, the system described in this paragraph. A 7-page typewritten letter, built up from page, line, and letter-number references to a book known only to the communicants, was intercepted by the British and turned over to the United States Government for use in connection with the prosecution of the Hindus for violating our neutrality. The author solved this message without the book in question, by taking full advantage of the clues referred to.

[8] Gyldén, Yves. *Chifferbydernas Insatser I Världskriget Till Lands*, Stockholm, 1931. A translation under the title *The Contribution of the Cryptographic Bureaus in the World War*, appeared in the Signal Corps Bulletin in seven successive installments, from November–December 1933 to November–December 1934, inclusive.

Nikolaieff, A. M. *Secret Causes of German success on the Eastern Front.* Coast Artillery Journal, September–October, 1935.

50. Analytical key for cryptanalysis.—*a.* It may be of assistance to indicate, by means of an outline, the relationships existing among the various cryptographic systems thus far considered. This graphic outline will be augmented from time to time as the different cipher systems are examined, and will constitute what has already been alluded to in Par. 6*d* and there termed an analytical key for cryptanalysis.[9] Fundamentally its nature is that of a schematic classification of the different systems examined. The analytical key forms an insert at the end of the book.

b. Note, in the analytical key, the rather clear-cut, dichotomous method of treatment; that is, classification by subdivision into pairs. For example, in the very first step there are only two alternatives: the cryptogram is either (1) cipher, or (2) code. If it is cipher, it is either (1) substitution, (2) transposition. If it is a substitution cipher, it is either (1) monographic, or (2) polygraphic—and so on. If the student will study the analytical key attentively, it will assist him in fixing in mind the manner in which the various systems covered thus far are related to one another, and this will be of benefit in clearing away some of the mental fog or haziness from which he is at first apt to suffer.

c. The numbers in parentheses refer to specific paragraphs in this text, so that the student may readily turn to the text for detailed information or for purposes of refreshing his memory as to procedure.

d. In addition to these reference numbers there have been affixed to the successive steps in the dichotomy, numbers that mark the "routes" on the cryptanalytic map (the analytical key) which the student cryptanalyst should follow if he wishes to facilitate his travels along the rather complicated and difficult road to success in cryptanalysis, in somewhat the same way in which an intelligent motorist follows the routes indicated on a geographical map if he wishes to facilitate his travels along unfamiliar roads. The analogy is only partially valid, however. The motorist usually knows in advance the distant point which he desires to reach and he proceeds thereto by the best and shortest route, which he finds by observing the route indications on a map and following the route markers on the road. Occasionally he encounters a detour but these are unexpected difficulties as a rule. Least of all does he anticipate any necessity for journeys down what may soon turn out to be blind alleys and "dead-end" streets, forcing him to double back on his way. Now the cryptanalyst also has a distant goal in mind—the solution of the cryptogram at hand—but he does not know at the outset of his journey the exact spot where it is located on the cryptanalytic map. The map contains many routes and he proceeds to test them one by one, in a successive chain. He encounters many blind alleys and dead-end streets, which force him to retrace his steps; he makes many detours and jumps many hurdles. Some of these retracings of steps, doubling back on his tracks, jumping of hurdles, and detours are unavoidable, but a few are avoidable. If properly employed, the analytical key will help the careful student to avoid those which should and can be avoided; if it does that much it will serve the principal purpose for which it is intended.

---

[9] This analytical key is quite analogous to the analytical keys usually found in the handbooks biologists commonly employ in the classification and identification of living organisms. In fact, there are several points of resemblance between, for example, that branch of biology called taxonomic botany and cryptanalysis. In the former the first steps in the classificatory process are based upon observation of externally quite marked differences; as the process continues, the observational details become finer and finer, involving more and more difficulties as the work progresses. Towards the end of the work the botanical taxonomist may have to dissect the specimen and study internal characteristics. The whole process is largely a matter of painstaking, accurate observation of data and drawing proper conclusions therefrom. Except for the fact that the botanical taxonomist depends almost entirely upon ocular observation of characteristics while the cryptanalyst in addition to observation must use some statistics, the steps taken by the former are quite similar to those taken by the latter. It is only at the very end of the work that a significant dissimilarity between the two sciences arises. If the botanist makes a mistake in observation or deduction, he merely fails to identify the specimen correctly; he has an "answer"—but the answer is wrong. He may not be cognizant of the error; however, other more skillful botanists will find him out. But if the cryptanalyst makes a mistake in observation or deduction, he fails to get any "answer" at all; he needs nobody to tell him he has failed. Further, there is one additional important point of difference. The botanist is studying a bit of Nature—and she does not consciously interpose obstacles, pitfalls, and dissimulations in the path of those trying to solve her mysteries. The cryptanalyst, on the other hand, is studying a piece of writing prepared with the express purpose of preventing its being read by any persons for whom it is not intended. The obstacles, pitfalls, and dissimulations are here consciously interposed by the one who cryptographed the message. These, of course, are what make cryptanalysis different and difficult.

e. The analytical key may, however, serve another purpose of a somewhat different nature. When a multitude of cryptographic systems of diverse types must be filed in some systematic manner apart from the names of the correspondents or other reference data, or if in conducting instructional activities classificatory designations are desirable, the reference numbers on the analytical key may be made to serve as "type numbers." Thus, instead of stating that a given cryptogram is a keyword-systematically-mixed-uniliteral-monoalphabetic-monographic substitution cipher one may say that it is a "Type 901 cryptogram."

f. The method of assigning type numbers is quite simple. If the student will examine the numbers he will note that successive levels in the dichotomy are designated by successive hundreds. Thus, the first level, the classification into cipher and code is assigned the numbers 101 and 102. On the second level, under cipher, the classification into monographic and polygraphic systems is assigned the numbers 201 and 202, etc. Numbers in the same hundreds apply therefore to systems at the same level in the classification. There is no particular virtue in this scheme of assigning type numbers except that it provides for a considerable degree of expansion in future studies.

APPENDIX 1

APPENDIX

Table No. Page

1–A. Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters. Arranged alphabetically 108

1–B. Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters. Arranged according to frequency 109

1–C. Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters 110

2–A. Absolute frequency of letters appearing in the combined five sets of messages totalling 50,000 letters. Arranged alphabetically 109

2–B. Absolute frequency of letters appearing in the combined five sets of messages totalling 50,000 letters. Arranged according to frequency 110

2–C. Absolute frequency of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in the combined five sets of messages totalling 50,000 letters 110

2–D. Absolute frequencies of letters as initial letters of 10,000 words found in Government plain-text telegrams. (1) Arranged alphabetically, and (2) according to absolute frequencies 111

2–E. Absolute frequencies of letters as final letters of 10,000 words found in Government plain-text telegrams. (1) Arranged alphabetically, and (2) according to absolute frequencies 111

3. Relative frequencies of letters appearing in 1,000 letters based upon Table 2. (1) Arranged alphabetically, (2) according to absolute frequency, (3) vowels, (4) high frequency consonants, (5) medium frequency consonants, and (6) low frequency consonants 112

4. Frequency distribution for 10,000 letters of literary English. (1) Arranged alphabetically, and (2) according to absolute frequencies 113

5. Frequency distribution for 10,000 letters of telegraphic English. (1) Arranged alphabetically, and (2) according to absolute frequencies 113

6. Frequency distribution of digraphs, based on 50,000 letters of Government plain-text telegrams, reduced to 5,000 digraphs 113

7–A. The 428 different digraphs of Table 6. Arranged according to their absolute frequencies 114

7–B. The 18 digraphs composing 25% of the digraphs in Table 6. Arranged alphabetically according to their initial letters, (1) and according to their final letters (2) and according to their absolute frequencies 116

7–C. The 53 digraphs composing 50% of the digraphs in Table 6. Arranged alphabetically according to their initial letters, (1) and according to their final letters (2) and according to their absolute frequencies 117

7–D. The 117 digraphs composing 75% of the digraphs in Table 6. Arranged alphabetically according to their initial letters, (1) and according to their final letters (2) and according to their absolute frequencies 118

7–E. All the 428 digraphs of Table 6. Arranged first alphabetically according to their initial letters and then alphabetically according to their final letters 119

(See Table 6. Read across the rows) 113

8. The 428 different digraphs of Table 6. Arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter 120

9–A. The 428 different digraphs of Table 6. Arranged first alphabetically according to their final letters and then according to their absolute frequencies 123

9–B. The 18 digraphs composing 25% of the 5,000 digraphs of Table 6. Arranged alphabetically according to their final letters, (1) and according to their initial letters, (2) and according to their absolute frequencies 126

9–C. The 53 digraphs composing 50% of the 5,000 digraphs of Table 6. Arranged alphabetically according to their final letters, (1) and according to their initial letters, (2) and according to their absolute frequencies 126

9–D. The 117 digraphs composing 75% of the 5,000 digraphs of Table 6. Arranged alphabetically according to their final letters, (1) and according to their initial letters, (2) and according to their absolute frequencies 127

9–E. All the 428 different digraphs of Table 6. Arranged alphabetically first according to their final letters and then according to their initial letters 129

(See Table 6. Read down the columns) 113

Table No. Page
10– The 56 trigraphs appearing 100 or more times in the 50,000 letters of government plain-text telegrams—
–A. Arranged according to their absolute frequencies 129
–B. Arranged first alphabetically according to their initial letters and then according to their absolute frequencies 130
–C. Arranged first alphabetically according to their central letters and then according to their absolute frequencies 130
–D. Arranged first alphabetically according to their final letters and then according to their absolute frequencies 131
11– The 54 tetragraphs appearing 50 or more times in the 50,000 letters of government plain-text telegrams—
–A. Arranged according to their absolute frequencies 132
–B. Arranged first alphabetically according to their initial letters and then according to their absolute frequencies 132
–C. Arranged first alphabetically according to their second letters and then according to their absolute frequencies 133
–D. Arranged first alphabetically according to their third letters and then according to their absolute frequencies 133
–E. Arranged first alphabetically according to their final letters and then according to their absolute frequencies 134
12. Average and mean lengths of words 135

TABLE 1–A.—*Absolute frequencies of letters appearing in five sets of Governmental plain-text telegrams, each set containing 10,000 letters, arranged alphabetically*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1,367	E	1,294	E	1,292	E	1,270	E	1,275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	233	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10,000		10,000		10,000		10,000		10,000

TABLE 2–A.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters, arranged alphabetically*

A	3, 683	G	819	L	1, 821	Q	175	V	766
B	487	H	1, 694	M	1, 237	R	3, 788	W	780
C	1, 534	I	3, 676	N	3, 975	S	3, 058	X	231
D	2, 122	J	82	O	3, 764	T	4, 595	Y	967
E	6, 498	K	148	P	1, 335	U	1, 300	Z	49
F	1, 416								

TABLE 1–B.—*Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters, arranged according to frequency*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
T	936	T	879	T	894	T	958	T	928
N	786	N	794	N	815	N	800	R	786
R	760	A	783	O	791	O	756	N	780
I	742	O	770	I	787	A	740	O	762
A	738	I	750	R	762	R	735	A	741
O	685	R	745	A	681	I	700	I	697
S	658	S	583	S	585	S	628	S	604
D	387	D	413	D	423	D	451	D	448
L	365	L	393	H	335	L	386	H	349
C	319	H	351	L	333	H	349	L	344
H	310	C	300	P	317	C	326	C	301
U	270	F	287	U	312	F	287	F	281
F	253	P	272	F	308	M	249	M	268
M	242	M	240	C	288	U	247	P	260
P	241	U	233	M	238	P	245	U	238
Y	191	G	175	Y	179	Y	213	Y	229
G	166	V	173	G	161	G	167	W	182
W	166	W	163	V	142	V	133	V	155
V	163	Y	155	W	136	W	133	G	150
B	104	B	103	B	98	B	83	B	99
X	43	X	50	Q	45	X	53	X	41
Q	40	K	38	X	44	Q	38	K	31
K	36	Q	22	K	22	K	21	Q	30
J	18	J	17	J	10	J	21	J	16
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

Table 1–C.—*Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters*

Set No.	Vowels	High Frequency Consonants	Medium Frequency Consonants	Low Frequency Consonants
1	3, 993	3, 527	2, 329	151
2	3, 985	3, 414	2, 457	144
3	4, 042	3, 479	2, 356	123
4	3, 926	3, 572	2, 358	144
5	3, 942	3, 546	2, 389	123
Total [1]	19, 888	17, 538	11, 889	685

[1] Grand total, 50,000.

Table 2–B.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters arranged according to frequencies*

E	6, 498	I	3, 676	C	1, 534	Y	967	X	231
T	4, 595	S	3, 058	F	1, 416	G	819	Q	175
N	3, 975	D	2, 122	P	1, 335	W	780	K	148
R	3, 788	L	1, 821	U	1, 300	V	766	J	82
O	3, 764	H	1, 694	M	1, 237	B	487	Z	49
A	3, 683								

Table 2–C.—*Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in the combined five sets of messages totalling 50,000 letters*

Vowels	19, 888
High Frequency Consonants (D, N, R, S, and T)	17, 538
Medium Frequency Consonants (B, C, F, G, H, L, M, P, V, and W)	11, 889
Low Frequency Consonants (J, K, Q, X, and Z)	685
Total	50, 000

TABLE 2-D.—*Absolute frequencies of letters as initial letters of 10,000 words found in Government plain-text telegrams*

(1) ARRANGED ALPHABETICALLY

A	905	G	109	L	196	Q	30	V	77
B	287	H	272	M	384	R	611	W	320
C	664	I	344	N	441	S	965	X	4
D	525	J	44	O	646	T	1,253	Y	88
E	390	K	23	P	433	U	122	Z	12
F	855								
								Total	10,000

(2) ARRANGED ACCORDING TO ABSOLUTE FREQUENCIES

T	1,253	R	611	M	384	L	196	J	44
S	965	D	525	I	344	U	122	Q	30
A	905	N	441	W	320	G	109	K	23
F	855	P	433	B	287	Y	88	Z	12
C	664	E	390	H	272	V	77	X	4
O	646								
								Total	10,000

TABLE 2-E.—*Absolute frequencies of letters as final letters of 10,000 words found in Government plain-text telegrams*

(1) ARRANGED ALPHABETICALLY

A	269	G	225	L	354	Q	8	V	4
B	22	H	450	M	154	R	769	W	45
C	86	I	22	N	872	S	962	X	116
D	1,002	J	6	O	575	T	1,007	Y	866
E	1,628	K	53	P	213	U	31	Z	9
F	252								
								Total	10,000

(2) ARRANGED ACCORDING TO ABSOLUTE FREQUENCIES

E	1,628	R	769	F	252	C	86	I	22
T	1,007	O	575	G	225	K	53	Z	9
D	1,002	H	450	P	213	W	45	Q	8
S	962	L	354	M	154	U	31	J	6
N	872	A	269	X	116	B	22	V	4
Y	866								
								Total	10,000

TABLE 3.—*Relative frequencies of letters appearing in 1,000 letters based upon Table 2-B*

(1) ARRANGED ALPHABETICALLY

A	73.66	G	16.38	L	36.42	Q	3.50	V	15.32
B	9.74	H	33.88	M	24.74	R	75.76	W	15.60
C	30.68	I	73.52	N	79.50	S	61.16	X	4.62
D	42.44	J	1.64	O	75.28	T	91.90	Y	19.34
E	129.96	K	2.96	P	26.70	U	26.00	Z	.98
F	28.32								
								Total	1,000.00

(2) ARRANGED ACCORDING TO FREQUENCY

E	129.96	I	73.52	C	30.68	Y	19.34	X	4.62
T	91.90	S	61.16	F	28.32	G	16.38	Q	3.50
N	79.50	D	42.44	P	26.70	W	15.60	K	2.96
R	75.76	L	36.42	U	26.00	V	15.32	J	1.64
O	75.28	H	33.88	M	24.74	B	9.74	Z	.98
A	73.66								
								Total	1,000.00

(3) VOWELS

A	73.66
E	129.96
I	73.52
O	75.28
U	26.00
Y	19.34
Total	397.76

(4) HIGH-FREQUENCY CONSONANTS

D	42.44
N	79.50
R	75.76
S	61.16
T	91.90
Total	350.76

(5) MEDIUM-FREQUENCY CONSONANTS

B	9.74
C	30.68
F	28.32
G	16.38
H	33.88
L	36.42
M	24.74
P	26.70
V	15.32
W	15.60
Total	237.78

(6) LOW-FREQUENCY CONSONANTS

X	4.62
Q	3.50
K	2.96
J	1.64
Z	.98
Total	13.70
Total (3), (4), (5), (6)	1,000.00

TABLE 6.—*Frequency distribution of digraphs—Based on 50,000 letters of Government plain-text telegrams; reduced to 5,000 digraphs*

	Second Letter																											
First Letter	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	Total	Blanks
A	3	6	14	27	1	4	6	2	17	1	2	32	14	64	2	12		44	41	47	13	7	3		12		374	3
B	4				18				2	1		6	1		4			2	1	1	2				7		49	14
C	20		3	1	32	1		14	7		4	5	1	1	41			4	1	14	4		1		1		155	8
D	32	4	4	8	33	8	2	2	27	1		3	5	4	16	5	2	12	13	15	5	3	4		1		209	3
E	35	4	32	60	42	18	4	7	27	1		29	14	111	12	20	12	87	54	37	3	20	7	7	4	1	648	1
F	5		2	1	10	11	1		39			2	1		40	1		9	3	11	3		1		1		141	9
G	7		2	1	14	2	1	20	5	1		2	1	3	6	2		5	3	4	2		1				82	7
H	20	1	3	2	20	5			33			1	2	3	20	1	1	17	4	28	8		1		1		171	7
I	8	2	22	6	13	10	19				2	23	9	75	41	7		27	35	27		25		15		2	368	7
J	1				2										2						2						7	22
K	1		1		6				2			1		1					1								13	19
L	28	3	3	9	37	3	1	1	20			27	2	1	13	3		2	6	8	2	2	2		10		183	5
M	36	6	3	1	26	1		1	9				13		10	8		2	4	2	2				2		126	10
N	26	2	19	52	57	9	27	4	30	1	2	5	5	8	18	3	1	4	24	82	7	3	3		5		397	2
O	7	4	8	12	3	25	2	3	5	1	2	19	25	77	6	25		64	14	19	37	7	8	1	2		376	2
P	14	1	1	1	23	2		3	6			13	4	1	17	11		18	6	8	3	1	1		1		135	6
Q													1					1			15						17	23
R	39	2	9	17	98	6	7	3	30	1	1	5	9	7	28	13		11	31	42	5	5	4		9		382	3
S	24	3	13	5	49	12	2	26	34		1	2	3	4	15	10		5	19	63	11	1	4		1		307	4
T	28	3	6	6	71	7	1	78	45			5	6	7	50	2	1	17	19	19	5		36		41	1	454	4
U	5	3	3	3	11	1	8		5			6	5	21	1	2		31	12	12		1					130	9
V	6				57				12						1					1							77	21
W	12				22			4	13			1		2	19			1	1						1		76	16
X	2		2	1	1	1		1	2					1	1	2		1	1	7							23	13
Y	6	2	4	4	9	11	1	1	3			2	2	6	10	3		4	11	15	1		1				96	7
Z	1				2				1																		4	23
Total	370	46	154	217	657	137	82	170	374	8	14	189	123	397	373	130	17	368	304	462	130	75	77	23	99	4	5,000	
Blanks	1	11	6	7	1	7	12	10	3	18	19	6	6	7	3	8	21	4	4	5	7	15	11	23	10	23		248

TABLE 4.—*Frequency distribution for 10,000 letters of literary English, as compiled by Hitt* [1]

(1) ALPHABETICALLY ARRANGED

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

TABLE 5.—*Frequency distribution for 10,000 letters of telegraphic English as compiled by Hitt*

(1) ALPHABETICALLY ARRANGED

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

---

[1] Hitt, Capt. Parker. *Manual for the Solution of Military Ciphers.*

TABLE 7-A.—*The 428 different digraphs of table 6 arranged according to their absolute frequencies*

EN	111
RE	98
ER	87
NT	82
TH	78
ON	77
IN	75
TE	71
AN	64
OR	64
ST	63
ED	60
NE	57
VE	57
ES	54
ND	52
TO	50
SE	49
	[1]1,249
AT	47
TI	45
AR	44
EE	42
RT	42
AS	41
CO	41
IO	41
TY	41
FO	40
FI	39
RA	39
ET	37
OU	37
LE	37
MA	36
TW	36
EA	35
IS	35
SI	34
DE	33
HI	33
AL	32
CE	32
DA	32
EC	32
RS	31
UR	31
NI	30
RI	30
EL	29
HT	28
LA	28
RO	28
TA	28
	[2]2,495
LL	27
AD	27
DI	27
EI	27
IR	27
IT	27
NG	27
ME	26
NA	26
SH	26
IV	25
OF	25
OM	25
OP	25
NS	24
SA	24
IL	23
PE	23
IC	22
WE	22
UN	21
CA	20
EP	20
EV	20
GH	20
HA	20
HE	20
HO	20
LI	20
SS	19
TT	19
IG	19
NC	19
OL	19
OT	19
TS	19
WO	19
BE	18
EF	18
NO	18
PR	18
AI	17
HR	17
PO	17
RD	17
TR	17
DO	16
DT	15
IX	15
QU	15
SO	15
YT	15
AC	14
AM	14
CH	14
CT	14
EM	14
GE	14
OS	14
PA	14
PL	13
RP	13
SC	13
WI	13
MM	13
DS	13
AU	13
IE	13
LO	13
	[3]3,745
AP	12
DR	12
EQ	12
AY	12
EO	12
OD	12
SF	12
US	12
UT	12
VI	12
WA	12
FF	11
PP	11
RR	11
UE	11
FT	11
SU	11
YF	11
YS	11
YO	10
FE	10
IF	10
LY	10
MO	10
SP	10
YE	9
FR	9
IM	9
LD	9
MI	9
NF	9
RC	9
RM	9
RY	9
DD	8
NN	8
DF	8
IA	8
HU	8
LT	8
MP	8
OC	8
OW	8
PT	8
UG	8
AV	7
BY	7
CI	7
EH	7
OA	7
EW	7
EX	7

---

[1] The 18 digraphs above this line compose 25% of the total.
[2] The 53 digraphs above this line compose 50% of the total.
[3] The 117 digraphs above this line compose 75% of the total.

TABLE 7-A.—*The 428 different digraphs of table 6 arranged according to their absolute frequencies*—Continued

GA	7	SD	5	DV	3	KI	2
IP	7	SR	5	AA	3	LM	2
NU	7	TL	5	EU	3	LR	2
OV	7	TU	5	OE	3	LU	2
RG	7	UM	5	YI	3	LV	2
RN	7	AF	4	FS	3	LW	2
TE	7	BA	4	FU	3	MR	2
TN	7	BO	4	GN	3	MT	2
XT	7	CK	4	GS	3	MU	2
AB	6	CR	4	HC	3	MY	2
AG	6	CU	4	HN	3	NB	2
BL	6	DB	4	LB	3	NK	2
OO	6	DC	4	LC	3	OG	2
YA	6	DN	4	LF	3	OK	2
GO	6	DW	4	LP	3	PF	2
ID	6	EB	4	MC	3	RB	2
KE	6	EG	4	NP	3	SG	2
LS	6	EY	4	NV	3	SL	2
MB	6	GT	4	NW	3	TP	2
PI	6	HS	4	OH	3	UP	2
PS	6	MS	4	AH	2	WN	2
RF	6	NH	4	AK	2	XA	2
TC	6	NR	4	BI	2	XC	2
TD	6	OB	4	BR	2	XI	2
TM	6	PM	4	BU	2	XP	2
UL	6	RW	4	DG	2	YB	2
VA	6	SN	4	DH	2	YL	2
YN	6	SW	4	DO	2	YM	2
CL	5	WH	4	AO	2	ZE	2
DM	5	YC	4	OY	2	GG	1
DP	5	YD	4	FC	2	AJ	1
DU	5	YR	4	FL	2	BJ	1
OI	5	PH	3	GC	2	BM	1
UA	5	PU	3	GF	2	BS	1
UI	5	RH	3	GL	2	BT	1
FA	5	SB	3	GP	2	CD	1
GI	5	SM	3	GU	2	CF	1
GR	5	TB	3	HD	2	CM	1
HF	5	UB	3	HM	2	CN	1
NL	5	UC	3	IB	2	CS	1
NM	5	UD	3	IK	2	CW	1
NY	5	YP	3	IZ	2	CY	1
RL	5	CC	3	JE	2	DJ	1
RU	5	AW	3	JO	2	DY	1
RV	5	DL	3	JU	2	EJ	1

TABLE 7-A.—*The 428 different digraphs of table 6 arranged according to their absolute frequencies*—Continued

AE	1	HY	1	PD	1	WL	1
UO	1	JA	1	PN	1	WR	1
YU	1	KA	1	PV	1	WS	1
EZ	1	KC	1	PW	1	WY	1
FD	1	KL	1	PY	1	XD	1
FG	1	KN	1	QM	1	XE	1
FM	1	KS	1	QR	1	XF	1
FP	1	LG	1	RJ	1	XH	1
FW	1	LH	1	RK	1	XN	1
FY	1	LN	1	SK	1	XO	1
GD	1	MD	1	SV	1	XR	1
GJ	1	MF	1	SY	1	XS	1
GM	1	MH	1	TG	1	YG	1
GW	1	NJ	1	TQ	1	YH	1
HB	1	NQ	1	TZ	1	YW	1
HL	1	OJ	1	UF	1	ZA	1
HP	1	OX	1	UV	1	ZI	1
HQ	1	PB	1	VO	1		
HW	1	PC	1	VT	1	Total	5,000

TABLE 7-B.—*The 18 digraphs composing 25% of the digraphs in Table 6 arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AN	64	ON	77
		OR	64
ED	60	RE	98
EN	111		
ER	87	SE	49
ES	54	ST	63
		TE	71
IN	75	TH	78
		TO	50
ND	52	VE	57
NE	57		
NT	82	Total	1,249

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64	ON	77
		OR	64
EN	111	RE	98
ER	87		
ED	60	SE	49
ES	54	ST	63
		TH	78
IN	75	TE	71
		TO	50
NT	82	VE	57
NE	57		
ND	52	Total	1,249

TABLE 7–C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AL	32
AN	64
AR	44
AS	41
AT	47
CE	32
CO	41
DA	32
DE	33
EA	35
EC	32
ED	60
EE	42
EL	29
EN	111
ER	87
ES	54
ET	37
FI	39
FO	40
HI	33
HT	28
IN	75
IO	41
IS	35
LA	28
LE	37
MA	36
ND	52
NE	57
NI	30
NT	82
ON	77
OR	64
OU	37
RA	39
RE	98
RI	30
RO	28
RS	31
RT	42
SE	49
SI	34
ST	63
TA	28
TE	71
TH	78
TI	45
TO	50
TW	36
TY	41
UR	31
VE	57
Total	2,495

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64
AT	47
AR	44
AS	41
AL	32
CO	41
CE	32
DE	33
DA	32
EN	111
ER	87
ED	60
ES	54
EE	42
ET	37
EA	35
EC	32
EL	29
FO	40
FI	39
HI	33
HT	28
IN	75
IO	41
IS	35
LE	37
LA	28
MA	36
NT	82
NE	57
ND	52
NI	30
ON	77
OR	64
OU	37
RE	98
RT	42
RA	39
RS	31
RI	30
RO	28
ST	63
SE	49
SI	34
TH	78
TE	71
TO	50
TI	45
TY	41
TW	36
TA	28
UR	31
VE	57
Total	2,495

TABLE 7-D.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters*——

(1) AND ACCORDING TO THEIR FINAL LETTERS

AC	14
AD	27
AI	17
AL	32
AM	14
AN	64
AR	44
AS	41
AT	47
AU	13
BE	18
CA	20
CE	32
CH	14
CO	41
CT	14
DA	32
DE	33
DI	27
DO	16
DS	13
DT	15
EA	35
EC	32
ED	60
EE	42
EF	18
EI	27
EL	29
EM	14
EN	111
EP	20
ER	87
ES	54
ET	37
EV	20
FI	39
FO	40
GE	14
GH	20
HA	20
HE	20
HI	33
HO	20
HR	17
HT	28
IC	22
IE	13
IG	19
IL	23
IN	75
IO	41
IR	27
IS	35
IT	27
IV	25
IX	15
LA	28
LE	37
LI	20
LL	27
LO	13
MA	36
ME	26
NA	26
NC	19
ND	52
NE	57
NG	27
NI	30
NO	18
NS	24
NT	82
OF	25
OL	19
OM	25
ON	77
OP	25
OR	64
OS	14
OT	19
OU	37
PA	14
PE	23
PO	17
PR	18
QU	15
RA	39
RD	17
RE	98
RI	30
RO	28
RS	31
RT	42
SA	24
SE	49
SH	26
SI	34
SO	15
SS	19
ST	63
TA	28
TE	71
TH	78
TI	45
TO	50
TR	17
TS	19
TT	19
TW	36
TY	41
UN	21
UR	31
VE	57
WE	22
WO	19
YT	15
Total	3,745

Table 7–D, Concluded.—*The 117 digraphs comprising 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters*——

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64
AT	47
AR	44
AS	41
AL	32
AD	27
AI	17
AC	14
AM	14
AU	13
BE	18
CO	41
CE	32
CA	20
CH	14
CT	14
DE	33
DA	32
DI	27
DO	16
DT	15
DS	13
EN	111
ER	87
ED	60
ES	54
EE	42
ET	37
EA	35
EC	32
EL	29
EI	27
EP	20
EV	20
EF	18
EM	14
FO	40
FI	39
GH	20
GE	14
HI	33
HT	28
HA	20
HE	20
HO	20
HR	17
IN	75
IO	41
IS	35
IR	27
IT	27
IV	25
IL	23
IC	22
IG	19
IX	15
IE	13
LE	37
LA	28
LL	27
LI	20
LO	13
MA	36
ME	26
NT	82
NE	57
ND	52
NI	30
NG	27
NA	26
NS	24
NC	19
NO	18
ON	77
OR	64
OU	37
OF	25
OM	25
OP	25
OL	19
OT	19
OS	14
PE	23
PR	18
PO	17
PA	14
QU	15
RE	98
RT	42
RA	39
RS	31
RI	30
RO	28
RD	17
ST	63
SE	49
SI	34
SH	26
SA	24
SS	19
SO	15
TH	78
TE	71
TO	50
TI	45
TY	41
TW	36
TA	28
TS	19
TT	19
TR	17
UR	31
UN	21
VE	57
WE	22
WO	19
YT	15
Total	3, 745

Table 7–E.—*All the 428 digraphs of Table 6, arranged first alphabetically according to their initial letters and then alphabetically according to their final letters.*

(SEE TABLE 6.—READ ACROSS THE ROWS)

TABLE 8.—*The 428 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

AN	64
AT	47
AR	44
AS	41
AL	32
AD	27
AI	17
AC	14
AM	14
AU	13
AP	12
AY	12
AV	7
AB	6
AG	6
AF	4
AA	3
AW	3
AH	2
AK	2
AO	2
AE	1
AJ	1
BE	18
BY	7
BL	6
BA	4
BO	4
BI	2
BR	2
BU	2
BJ	1
BM	1
BS	1
BT	1
CO	41
CE	32
CA	20
CH	14
CT	14
CI	7
CL	5
CK	4
CR	4
CU	4
CC	3
CD	1
CF	1
CM	1
CN	1
CS	1
CW	1
CY	1
DE	33
DA	32
DI	27
DO	16
DT	15
DS	13
DR	12
DD	8
DF	8
DM	5
DP	5
DU	5
DB	4
DC	4
DN	4
DW	4
DL	3
DV	3
DG	2
DH	2
DQ	2
DJ	1
DY	1
EN	111
ER	87
ED	60
ES	54
EE	42
ET	37
EA	35
EC	32
EL	29
EI	27
EP	20
EV	20
EF	18
EM	14
EO	12
EQ	12
EH	7
EW	7
EX	7
EB	4
EG	4
EY	4
EU	3
EJ	1
EZ	1
FO	40
FI	39
FF	11
FT	11
FE	10
FR	9
FA	5
FS	3
FU	3
FC	2
FL	2
FD	1
FG	1
FM	1
FP	1
FW	1
FY	1
GH	20
GE	14
GA	7
GO	6
GI	5
GR	5
GT	4
GN	3
GS	3
GC	2
GF	2
GL	2
GP	2
GU	2
GD	1
GG	1
GJ	1
GM	1
GW	1
HI	33
HT	28
HA	20
HE	20
HO	20
HR	17
HU	8
HF	5
HS	4
HC	3
HN	3
HD	2
HM	2
HB	1
HL	1
HP	1
HQ	1
HW	1
HY	1

---

[1] For arrangement alphabetically first under intial letters and then under final letters, see Table 6.

TABLE 8, Contd.—*The 428 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

IN	75
IO	41
IS	35
IR	27
IT	27
IV	25
IL	23
IC	22
IG	19
IX	15
IE	13
IF	10
IM	9
IA	8
IP	7
ID	6
IB	2
IK	2
IZ	2
JE	2
JO	2
JU	2
JA	1
KE	6
KI	2
KA	1
KC	1
KL	1
KN	1
KS	1
LE	37
LA	28
LL	27
LI	20
LO	13
LY	10
LD	9
LT	8
LS	6
LB	3
LC	3
LF	3
LP	3
LM	2
LR	2
LU	2
LV	2
LW	2
LG	1
LH	1
LN	1
MA	36
ME	26
MM	13
MO	10
MI	9
MP	8
MB	6
MS	4
MC	3
MR	2
MT	2
MU	2
MY	2
MD	1
MF	1
MH	1
NT	82
NE	57
ND	52
NI	30
NG	27
NA	26
NS	24
NC	19
NO	18
NF	9
NN	8
NU	7
NL	5
NM	5
NY	5
NH	4
NR	4
NP	3
NV	3
NW	3
NB	2
NK	2
NJ	1
NQ	1
ON	77
OR	64
OU	37
OF	25
OM	25
OP	25
OL	19
OT	19
OS	14
OD	12
OC	8
OW	8
OA	7
OV	7
OO	6
OI	5
OB	4
OE	3
OH	3
OG	2
OK	2
OY	2
OJ	1
OX	1
PE	23
PR	18
PO	17
PA	14
PL	13
PP	11
PT	8
PI	6
PS	6
PM	4
PH	3
PU	3
PF	2
PB	1
PC	1
PD	1
PN	1
PV	1
PW	1
PY	1
QU	15
QM	1
QR	1

---

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6.

TABLE 8, Concluded.—*The 428 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter*[1]

RE	98
RT	42
RA	39
RS	31
RI	30
RO	28
RD	17
RP	13
RR	11
RC	9
RM	9
RY	9
RG	7
RN	7
RF	6
RL	5
RU	5
RV	5
RW	4
RH	3
RB	2
RJ	1
RK	1
ST	63
SE	49
SI	34
SH	26
SA	24
SS	19
SO	15
SC	13
SF	12
SU	11
SP	10
SD	5
SR	5
SN	4
SW	4
SB	3
SM	3
SG	2
SL	2
SK	1
SV	1
SY	1
TH	78
TE	71
TO	50
TI	45
TY	41
TW	36
TA	28
TS	19
TT	19
TR	17
TF	7
TN	7
TC	6
TD	6
TM	6
TL	5
TU	5
TB	3
TP	2
TG	1
TQ	1
TZ	1
UR	31
UN	21
US	12
UT	12
UE	11
UG	8
UL	6
UA	5
UI	5
UM	5
UB	3
UC	3
UD	3
UP	2
UF	1
UO	1
UV	1
VE	57
VI	12
VA	6
VO	1
VT	1
WE	22
WO	19
WI	13
WA	12
WH	4
WN	2
WL	1
WR	1
WS	1
WY	1
XT	7
XA	2
XC	2
XI	2
XP	2
XD	1
XE	1
XF	1
XH	1
XN	1
XO	1
XR	1
XS	1
YT	15
YF	11
YS	11
YO	10
YE	9
YA	6
YN	6
YC	4
YD	4
YR	4
YI	3
YP	3
YB	2
YL	2
YM	2
YG	1
YH	1
YU	1
YW	1
ZE	2
ZA	1
ZI	1
Total	5,000

---

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6.

TABLE 9-A.—*The 428 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

RA	39
MA	36
EA	35
DA	32
LA	28
TA	28
NA	26
SA	24
CA	20
HA	20
PA	14
WA	12
IA	8
GA	7
OA	7
VA	6
YA	6
FA	5
UA	5
BA	4
AA	3
XA	2
JA	1
KA	1
ZA	1

AB	6
MB	6
DB	4
EB	4
OB	4
LB	3
SB	3
TB	3
UB	3
IB	2
NB	2
RB	2
YB	2
HB	1
PB	1

EC	32
IC	22
NC	19
AC	14
SC	13
RC	9
OC	8
TC	6
DC	4
YC	4
CC	3
HC	3
LC	3
MC	3
UC	3
FC	2
GC	2
XC	2
KC	1
PC	1

ED	60
ND	52
AD	27
RD	17
OD	12
LD	9
DD	8
ID	6
TD	6
SD	5
YD	4
UD	3
HD	2
CD	1
FD	1
GD	1
MD	1
PD	1
XD	1

RE	98
TE	71
NE	57
VE	57
SE	49
EE	42
LE	37
DE	33
CE	32
ME	26
PE	23
WE	22
HE	20
BE	18
GE	14
IE	13
UE	11
FE	10
YE	9
KE	6
OE	3
JE	2
ZE	2
AE	1
XE	1

OF	25
EF	18
SF	12
FF	11
YF	11
IF	10
NF	9
DF	8
TF	7
RF	6
HF	5
AF	4
LF	3
GF	2
PF	1
CF	2
MF	1
UF	1
XF	1

NG	27
IG	19
UG	8
RG	7
AG	6
EG	4
DG	2
OG	2
SG	2
FG	1
GG	1
LG	1
TG	1
YG	1

TH	78
SH	26
GH	20
CH	14
EH	7
NH	4
WH	4
OH	3
PH	3
RH	3
AH	2
DH	2
LH	1
MH	1
XH	1
YH	1

Table 9–A, Contd.—*The 428 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

TI	45
FI	39
SI	34
HI	33
NI	30
RI	30
DI	27
EI	27
LI	20
AI	17
WI	13
VI	12
MI	9
CI	7
PI	6
GI	5
OI	5
UI	5
YI	3
BI	2
KI	2
XI	2
ZI	1

AJ	1
BJ	1
DJ	1
EJ	1
GJ	1
NJ	1
OJ	1
RJ	1

CK	4
AK	2
IK	2
NK	2
OK	2
RK	1
SK	1

AL	32
EL	29
LL	27
IL	23
OL	19
PL	13
BL	6
UL	6
CL	5
NL	5
RL	5
TL	5
DL	3
FL	2
GL	2
SL	2
YL	2
HL	1
KL	1
WL	1

OM	25
AM	14
EM	14
MM	13
IM	9
RM	9
TM	6
DM	5
NM	5
UM	5
PM	4
SM	3
HM	2
LM	2
YM	2
BM	1
CM	1
FM	1
GM	1
QM	1

EN	111
ON	77
IN	75
AN	64
UN	21
NN	8
RN	7
TN	7
YN	6
DN	4
SN	4
GN	3
HN	3
WN	2
CN	1
KN	1
LN	1
PN	1
XN	1

TO	50
CO	41
IO	41
FO	40
RO	28
HO	20
WO	19
NO	18
PO	17
DO	16
SO	15
LO	13
EO	12
MO	10
YO	10
GO	6
OO	6
BO	4
AO	2
JO	2
UO	1
VO	1
XO	1

OP	25
EP	20
RP	13
AP	12
PP	11
SP	10
MP	8
IP	7
DP	5
LP	3
NP	3
YP	3
GP	2
TP	2
UP	2
XP	2
FP	1
HP	1

EQ	12
DQ	2
HQ	1
NQ	1
TQ	1

ER	87
OR	64
AR	44
UR	31
IR	27
PR	18
HR	17
TR	17
DR	12
RR	11
FR	9
GR	5
SR	5
CR	4
NR	4
YR	4
BR	2
LR	2
MR	2
QR	1
WR	1
XR	1

TABLE 9-A, Concluded.—*The 428 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

ES	54
AS	41
IS	35
RS	31
NS	24
SS	19
TS	19
OS	14
DS	13
US	12
YS	11
LS	6
PS	6
HS	4
MS	4
FS	3
GS	3
BS	1
CS	1
KS	1
WS	1
XS	1
NT	82
ST	63
AT	47
RT	42
ET	37
HT	28
IT	27
OT	19
TT	19
DT	15
YT	15
CT	14
UT	12
FT	11
LT	8
PT	8
XT	7
GT	4
MT	2
BT	1
VT	1
OU	37
QU	15
AU	13
SU	11
HU	8
NU	7
DU	5
RU	5
TU	5
CU	4
EU	3
FU	3
PU	3
BU	2
GU	2
JU	2
LU	2
MU	2
YU	1
IV	25
EV	20
AV	7
OV	7
RV	5
DV	3
NV	3
LV	2
PV	1
SV	1
UV	1
TW	36
OW	8
EW	7
DW	4
RW	4
SW	4
AW	3
NW	3
LW	2
CW	1
FW	1
GW	1
HW	1
PW	1
YW	1
IX	15
EX	7
OX	1
TY	41
AY	12
LY	10
RY	9
BY	7
NY	5
EY	4
MY	2
OY	2
CY	1
DY	1
FY	1
HY	1
PY	1
SY	1
WY	1
IZ	2
EZ	1
TZ	1
Total	5,000

TABLE 9–B.—*The 18 digraphs composing 25% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS

ED	60	IN	75
ND	52	ON	77
NE	57	TO	50
RE	98		
SE	49	ER	87
TE	71	OR	64
VE	57		
		ES	54
TH	78	NT	82
		ST	63
AN	64		
EN	111	Total	1, 249

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

ED	60	IN	75
ND	52	AN	64
RE	98	TO	50
TE	71		
NE	57	ER	87
VE	57	OR	64
SE	49		
		ES	54
TH	78	NT	82
		ST	63
EN	111		
ON	77	Total	1, 249

TABLE 9–C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS

DA	32	RE	98	EN	111	IS	35
EA	35	SE	49	IN	75	RS	31
LA	28	TE	71	ON	77		
MA	36	VE	57				
RA	39					AT	47
TA	28	TH	78	CO	41	ET	37
				FO	40	HT	28
				IO	41	NT	82
EC	32	FI	39	RO	28	RT	42
		HI	33	TO	50	ST	63
		NI	30				
ED	60	RI	30				
ND	52	SI	34	AR	44	OU	37
		TI	45	ER	87		
CE	32			OR	64	TW	36
DE	33	AL	32	UR	31		
EE	42	EL	29			TY	41
LE	37			AS	41		
NE	57	AN	64	ES	54	Total	2, 495

TABLE 9-C, Concluded.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

RA	39
MA	36
EA	35
DA	32
LA	28
TA	28
EC	32
ED	60
ND	52
RE	98
TE	71
NE	57
VE	57
SE	49
EE	42
LE	37
DE	33
CE	32
TH	78
TI	45
FI	39
SI	34
HI	33
NI	30
RI	30
AL	32
EL	29
EN	111
ON	77
IN	75
AN	64
TO	50
CO	41
IO	41
FO	40
RO	28
ER	87
OR	64
AR	44
UR	31
ES	54
AS	41
IS	35
RS	31
NT	82
ST	63
AT	47
RT	42
ET	37
HT	28
OU	37
TW	36
TY	41
Total	2,495

TABLE 9-D.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS

CA	20
DA	32
EA	35
HA	20
LA	28
MA	36
NA	26
PA	14
RA	39
SA	24
TA	28
AC	14
EC	32
IC	22
NC	19
AD	27
ED	60
ND	52
RD	17
BE	18
CE	32
DE	33
EE	42
GE	14
HE	20
IE	13
LE	37
ME	26
NE	57
PE	23
RE	98
SE	49
TE	71
VE	57
WE	22
EF	18
OF	25
IG	19
NG	27
CH	14
GH	20
SH	26
TH	78
AI	17
DI	27
EI	27
FI	39
HI	33
LI	20
NI	30
RI	30
SI	34
TI	45
AL	32
EL	29
IL	23
LL	27
OL	19
AM	14
EM	14
OM	25
AN	64
EN	111
IN	75
ON	77
UN	21

Table 9–D, Contd.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS—Continued

CO	41
DO	16
FO	40
HO	20
IO	41
LO	13
NO	18
PO	17
RO	28
SO	15
TO	50
WO	19
EP	20
OP	25
AR	44
TR	17
UR	31
ER	87
OR	64
PR	18
HR	17
IR	27
AS	41
SS	19
TS	19
DS	13
ES	54
NS	24
OS	14
IS	35
RS	31
AT	47
CT	14
DT	15
ET	37
HT	28
IT	27
NT	82
OT	19
RT	42
ST	63
TT	19
YT	15
AU	13
OU	37
QU	15
EV	20
IV	25
TW	36
IX	15
TY	41
Total	3, 745

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

RA	39
MA	36
EA	35
DA	32
LA	28
TA	28
NA	26
SA	24
CA	20
HA	20
PA	14
EC	32
IC	22
NC	19
AC	14
ED	60
ND	52
AD	27
RD	17
RE	98
TE	71
NE	57
VE	57
SE	49
EE	42
LE	37
DE	33
CE	32
ME	26
PE	23
WE	22
HE	20
BE	18
GE	14
IE	13
OF	25
EF	18
NG	27
IG	19
TH	78
SH	26
GH	20
CH	14
TI	45
FI	39
SI	34
HI	33
NI	30
RI	30
DI	27
EI	27
LI	20
AI	17
AL	32
EL	29
LL	27
IL	23
OL	19
OM	25
AM	14
EM	14
EN	111
ON	77
IN	75
AN	64
UN	21
TO	50
CO	41
IO	41
FO	40
RO	28
HO	20
WO	19
NO	18
PO	17
DO	16
SO	15
LO	13

TABLE 9–D, Concluded.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES—Continued

OP	25	ES	54	AT	47	QU	15
EP	20	AS	41	RT	42	AU	13
		IS	35	ET	37		
		RS	31	HT	28	IV	25
ER	87	NS	24	IT	27	EV	20
OR	64	SS	19	OT	19		
AR	44	TS	19	TT	19	TW	36
UR	31	OS	14	DT	15		
IR	27	DS	13	YT	15	IX	15
PR	18			CT	14	TY	41
HR	17	NT	82				
TR	17	ST	63	OU	37	Total	3,745

TABLE 9–E.—*All the 428 different digraphs of Table 6, arranged alphabetically first according to their final letters and then according to their initial letters*

(SEE TABLE 6.—READ DOWN THE COLUMNS)

TABLE 10–A.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged according to their absolute frequencies*

ENT	569	TOP	174	EIG	135
ION	260	NTH	171	FIV	135
AND	228	TWE	170	MEN	131
ING	226	TWO	163	SEV	131
IVE	225	ATI	160	ERS	126
TIO	221	THR	158	UND	125
FOR	218	NTY	157	NET	118
OUR	211	HRE	153	PER	115
THI	211	WEN	153	STA	115
ONE	210	FOU	152	TER	115
NIN	207	ORT	146	EQU	114
STO	202	REE	146	RED	113
EEN	196	SIX	146	TED	112
GHT	196	ASH	143	ERI	109
INE	192	DAS	140	HIR	106
VEN	190	IGH	140	IRT	105
EVE	177	ERE	138	DER	101
EST	176	COM	136	DRE	100
TEE	174	ATE	135		

TABLE 10-B. *The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their initial letters and then according to their absolute frequencies*

AND	228
ATI	160
ASH	143
ATE	135
COM	136
DAS	140
DER	101
DRE	100
ENT	569
EEN	196
EVE	177
EST	176
ERE	138
EIG	135
ERS	126
EQU	114
ERI	109
FOR	218
FOU	152
FIV	135
GHT	196
HRE	153
HIR	106
ION	260
ING	226
IVE	225
INE	192
IGH	140
IRT	105
MEN	131
NIN	207
NTH	171
NTY	157
NET	118
OUR	211
ONE	210
ORT	146
PER	115
REE	146
RED	113
STO	202
SIX	146
SEV	131
STA	115
TIO	221
THI	211
TEE	174
TOP	174
TWE	170
TWO	163
THR	158
TER	115
TED	112
UND	125
VEN	190
WEN	153

TABLE 10-C.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their central letters and then according to their absolute frequencies*

DAS	140
EEN	196
VEN	190
TEE	174
WEN	153
REE	146
MEN	131
SEV	131
NET	118
PER	115
TER	115
RED	113
TED	112
DER	101
IGH	140
THI	211
GHT	196
THR	158
TIO	221
NIN	207
SIX	146
EIG	135
FIV	135
HIR	106
ENT	569
AND	228
ING	226
ONE	210
INE	192
UND	125
ION	260
FOR	218
TOP	174
FOU	152
COM	136

Table 10–C, Concluded.—*The 56 trigrams appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their central letters and then according to their absolute frequencies*

EQU	114
HRE	153
ORT	146
ERE	138
ERS	126
ERI	109
IRT	105
DRE	100
EST	176
ASH	143
STO	202
NTH	171
ATI	160
NTY	157
ATE	135
STA	115
OUR	211
IVE	225
EVE	177
TWE	170
TWO	163

Table 10–D.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their final letters and then according to their absolute frequencies*

STA	115
AND	228
UND	125
RED	113
TED	112
IVE	225
ONE	210
INE	192
EVE	177
TEE	174
TWE	170
HRE	153
REE	146
ERE	138
ATE	135
DRE	100
ING	226
EIG	135
NTH	171
ASH	143
IGH	140
THI	211
ATI	160
ERI	109
COM	136
ION	260
NIN	207
EEN	196
VEN	190
WEN	153
MEN	131
TIO	221
STO	202
TWO	163
TOP	174
FOR	218
OUR	211
THR	158
PER	115
TER	115
HIR	106
DER	101
DAS	140
ERS	126
ENT	569
GHT	196
EST	176
ORT	146
NET	118
IRT	105
FOU	152
EQU	114
FIV	135
SEV	131
SIX	146
NTY	157

Table 11-A.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged according to their absolute frequencies*

TION	218
EVEN	168
TEEN	163
ENTY	161
STOP	154
WENT	153
NINE	153
TWEN	152
THRE	149
FOUR	144
IGHT	140
FIVE	135
HREE	134
EIGH	132
DASH	132
SEVE	121
ENTH	114
MENT	111
THIR	104
EENT	102
REQU	98
HIRT	97
COMM	93
QUES	87
UEST	87
EQUE	86
NDRE	77
OMMA	71
LLAR	71
OLLA	70
VENT	70
DOLL	68
LARS	68
THIS	68
PERI	67
ERIO	66
ASHT	64
HUND	64
DRED	63
RIOD	63
IVED	62
ENTS	62
FFIC	62
FROM	59
IRTY	59
RTEE	59
UNDR	59
NAUG	56
OURT	56
UGHT	56
STAT	54
AUGH	52
CENT	52
FICE	50

Table 11-B.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies*

ASHT	64
AUGH	52
COMM	93
CENT	52
DASH	132
DOLL	68
DRED	63
EVEN	168
ENTY	161
EIGH	132
ENTH	114
EENT	102
EQUE	86
ERIO	66
ENTS	62
FOUR	144
FIVE	135
FFIC	62
FROM	59
FICE	50
HREE	134
HIRT	97
HUND	64
IGHT	140
IVED	62
IRTY	59
LLAR	71
LARS	68
MENT	111
NINE	153
NDRE	77
NAUG	56
OMMA	71
OLLA	70
OURT	56
PERI	67
QUES	87
REQU	98
RIOD	63
RTEE	59
STOP	154
SEVE	121
STAT	54
TION	218
TEEN	163
TWEN	152
THRE	149
THIR	104
THIS	68
UEST	87
UNDR	59
UGHT	56
VENT	70
WENT	153

TABLE 11–C.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their second letters and then according to their absolute frequencies*

DASH	132
LARS	68
NAUG	56
NDRE	77
TEEN	163
WENT	153
SEVE	121
MENT	111
EENT	102
REQU	98
UEST	87
VENT	70
PERI	67
CENT	52
FFIC	62
IGHT	140
UGHT	56
THRE	149
THIR	104
THIS	68
TION	218
NINE	153
FIVE	135
EIGH	132
HIRT	97
RIOD	63
FICE	50
LLAR	71
OLLA	70
OMMA	71
ENTY	161
ENTH	114
ENTS	62
UNDR	59
FOUR	144
COMM	93
DOLL	68
EQUE	86
HREE	134
ERIO	66
DRED	63
FROM	59
IRTY	59
ASHT	64
STOP	154
RTEE	59
STAT	54
QUES	87
HUND	64
OURT	56
AUGH	52
EVEN	168
IVED	62
TWEN	152

TABLE 11–D.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their third letters and then according to their absolute frequencies*

LLAR	71
STAT	54
FICE	50
UNDR	59
EVEN	168
TEEN	163
TWEN	152
HREE	134
QUES	87
DRED	63
IVED	62
RTEE	59
EIGH	132
AUGH	52
IGHT	140
ASHT	64
UGHT	56
THIR	104
THIS	68
ERIO	66
FFIC	62
OLLA	70
DOLL	68
COMM	93
OMMA	71
WENT	153
NINE	153
MENT	111
EENT	102
VENT	70
HUND	64
CENT	52
TION	218
STOP	154
RIOD	63
FROM	59

TABLE 11-D, Concluded.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their third letters and then according to their absolute frequencies*

REQU	98
THRE	149
HIRT	97
NDRE	77
LARS	68
PERI	67
OURT	56
DASH	132
UEST	87
ENTY	161
ENTH	114
ENTS	62
IRTY	59
FOUR	144
EQUE	86
NAUG	56
FIVE	135
SEVE	121

TABLE 11-E.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their final letters and then according to their absolute frequencies*

OMMA	71
OLLA	70
FFIC	62
HUND	64
DRED	63
RIOD	63
IVED	62
NINE	153
THRE	149
FIVE	135
HREE	134
SEVE	121
EQUE	86
NDRE	77
RTEE	59
FICE	50
NAUG	56
DASH	132
EIGH	132
ENTH	114
AUGH	52
PERI	67
DOLL	68
COMM	93
FROM	59
TION	218
EVEN	168
TEEN	163
TWEN	152
ERIO	66
STOP	154
FOUR	144
THIR	104
LLAR	71
UNDR	59
QUES	87
THIS	68
LARS	68
ENTS	62
WENT	153
IGHT	140
MENT	111
EENT	102
HIRT	97
UEST	87
VENT	70
ASHT	64
UGHT	56
OURT	56
STAT	54
CENT	52
REQU	98
ENTY	161
IRTY	59

TABLE 12.—*Mean lengths of words*

Number of letters in word	Number of times word appears	Number of letters
1	378	378
2	973	1, 946
3	1, 307	3, 921
4	1, 635	6, 540
5	1, 410	7, 050
6	1, 143	6, 858
7	1, 009	7, 063
8	717	5, 736
9	476	4, 284
10	274	2, 740
11	161	1, 771
12	86	1, 032
13	23	299
14	23	322
15	4	60
120	9, 619	50, 000

(1) Average length of words ---- 7.5 Letters.
(2) Mean length of words ---- 5.2 Letters.
(3) Average length of messages ---- 217 Letters.
(4) Mean length of messages ---- 191 Letters.
(5) Mode (most frequent) length of messages ---- 105–114 Letters.
(6) It is extremely unusual to find 5 consecutive letters without at least one vowel.
(7) The average number of letters between vowels is 2.

INDEX

INDEX

	Paragraphs	Pages
Accented letters	5b	8.
Alphabets:		
Bipartite	35c	59.
Deciphering	31c	52.
Direct standard	12a, 16, 19	18, 26, 31-33.
Enciphering	29b, 31c	49, 52.
Keyword-mixed	31d	53.
Mixed	12a, 15a, 19, 21d, 22b, 24c, 31b.	18, 25, 31–33, 39, 39, 41, 52.
Reversed standard	12a. 16, 19b, 20b	18, 26, 33, 36.
Standard	12a, 15a, 16, 19, 20b, 23, 38e.	18, 25, 26, 31–33, 36, 40, 65.
Systematically mixed	31c, e	52, 53.
Analytic key for cryptanalysis	6d, 50	9, 103–104.
Arbitrary symbols	13h, 48	22, 100–101.
Assumptions	46h	98.
Average length of messages	11b	16.
Baconian cipher	35e	60.
Beginnings of messages	32e	54.
Biliteral substitution	41	70–71.
Bipartite alphabet	35b, c	59.
Blanks, number of	14e	24.
Book systems	49e	102
Censorship, methods for evading	47c	99.
Characteristic frequency of the letters of a language	9d, 14b, 25	12, 23, 41.
Characteristic frequency, suppression of	37, 41f	63, 71.
Checkerboard systems	44, 45	73–83, 83.
Checkerboards, 4-square	44	73–83.
Chinese Official Telegraph Code	48e	101.
Cipher:		
Baconian	35e	60.
Component	34	57–58.
Distinguished from code	6c, 38c	9, 64.
Text, length of, as compared with plain text	40c	69.
Unit	41c	70.
Classification of ciphers	12a, 13, 47, 50e, f	18, 18–22, 99, 104, 104.
Code systems	4a, 41g, 47b, 48d, e	7, 71, 99, 100, 101.
Distinguished from cipher	6c, 38c	9, 64.
Completing the plain component	20a, 34a	34, 57.
Concealed messages	47c	99.
Condensed table of repetitions	27i	46.
Consonants:		
Distinguished from vowels	28, 32c	46–47, 53.
Relative frequency of	10a, 13, 19	13, 18–22, 31–33.
In succession	32c	53.
Conversion of cipher text	21a, c, 34c	38, 38, 58.
Coordinates on work sheet	26d	42.

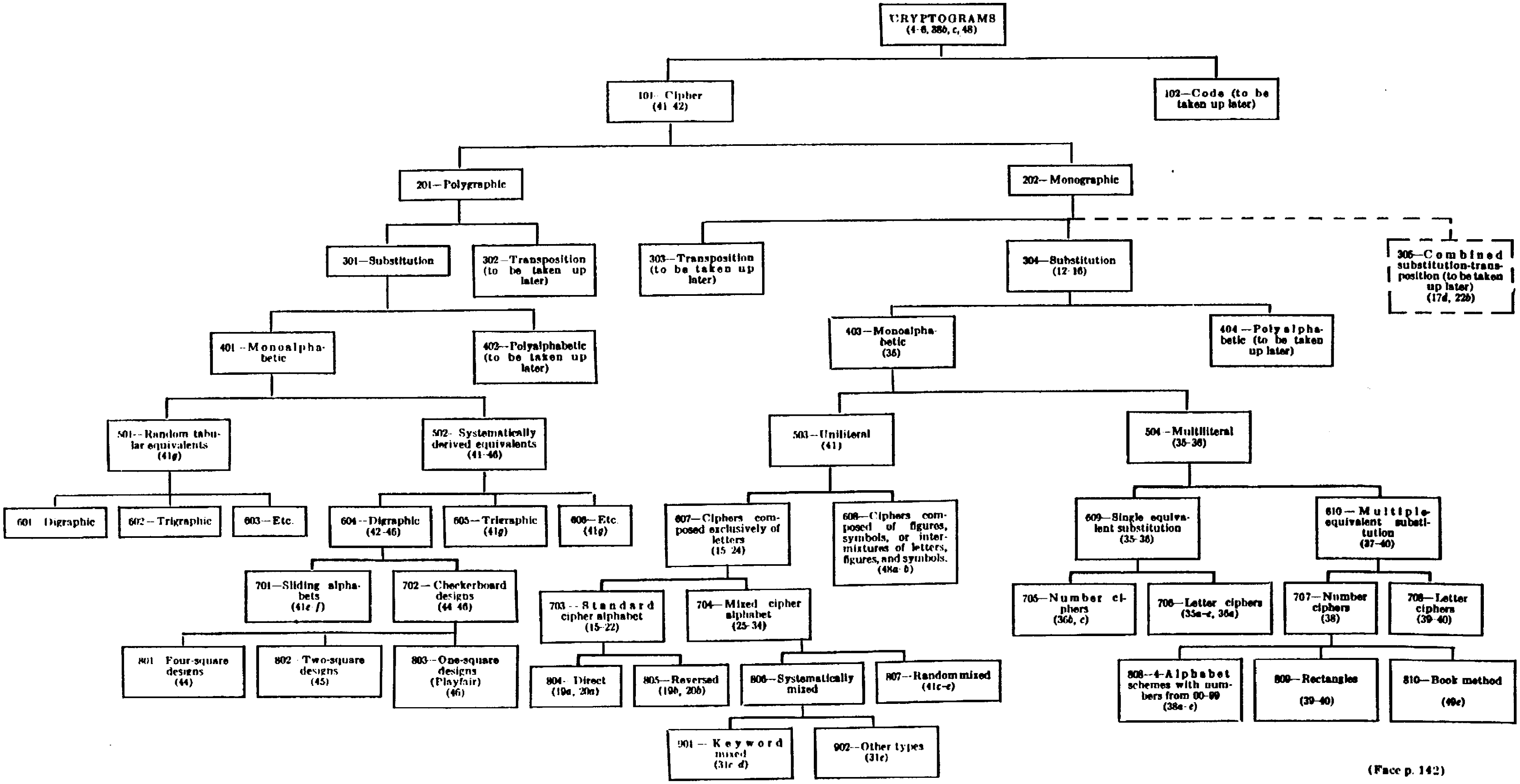
	Paragraphs	Pages
Coordination of services	2e	4.
Crests and troughs	10a, 14b, 41f, 44c	13, 23, 71, 74.
Absence of	14c	24.
Deciphering alphabets	31c	52.
Dictionary words used as code words	47b	99.
Digraphic substitution	41c, 42, 43	70, 72, 72–73.
Digraphs:		
Characteristic frequency	25	41.
Weighted according to relative frequency	29	48–49.
Distribution:		
Frequency	9a, 11b	11, 16.
Normal	17b, c	27, 28.
With no crests and troughs	14c	24
Dummy letters	47c	99.
Elementary sounds, characteristic frequency of	14b	23.
Enciphering alphabet	31c	52.
Endings of messages	32e	54.
Equivalent values	39b	66.
Figure ciphers	13h, 48	22, 100–101.
Fitting distribution to normal	17b, c, 19, 38e	27, 28, 31–33, 65.
Foreign language cryptograms	5b, c	8.
Formulas	33d	56.
Frequency distribution	9, 17, 19, 26e, 44c	11–13, 27–28, 31–33, 42, 74.
Fitted to normal	17b	27, 28.
For certain types of code	47b	99.
Four part	38d	65.
Multiliteral	35, 37, 41c	59–60, 63, 70.
Uniliteral	9, 17	11–13, 27–28.
Triliteral	27	43–46.
Frequency method of solution	18, 24d, 29	29–31, 41, 48–49.
General solutions in cryptanalysis	46i	98.
General system, determination of	4a, 6, 13, 50	7, 8–9, 18–22, 103–104.
Generatrix	20a	34.
Goodness of fit	17b	27.
Grilles	47c	99.
Hidden messages	47c	99.
High-frequency consonants	13d	19.
Historical examples of multiliteral systems	36	61.
Idiomorphism	33e	56.
Indicators	49b	101.
Intelligence facilities	2e	4.
Intelligible text obtained by chance	21b	38.
Intuitive method	33	55–57.
Invisible writing	1d	1.
Japanese Morse alphabet	5b, 48e	8, 101.
Kata Kana Morse alphabet	48e	101.
Key, analytical	6d, 50	9, 103–104.
Known sequences	23	40.
Language employed in a cryptogram	4a, 5	7, 8.
Language frequency characteristics	9d, 25	12, 41.
Language peculiarities	5b	8.

	Paragraphs	Pages
Letters:		
Accented	5b	8.
Low-frequency	31*c*	52.
Missing	5*b*, 14*e*	8, 24.
Low-frequency consonants	13*d*, 31*c*	19, 52.
Medium-frequency consonants	13*d*	19.
Messages:		
Beginnings and endings amenable to cryptanalysis	32*e*	54.
General phraseology	49*a*	101.
Hidden	47*c*	99.
Military text	10*b*	15.
Missing letters	5*b*, 14*e*	8, 24.
Mixed alphabet	15*a*, 22*b*, 24*c*, 81*b*	25, 39. 41, 52.
Mixed sequence	21*d*	39.
Modified Playfair	46*d*	86.
Monoalphabets	1*b*	1.
Monoalphabet distinguished from polyalphabet	12, 14	18, 22–25.
Morse alphabet, Japanese, Russian	5*b*, 48*e*	8, 101.
Multiliteral substitution	35, 37, 41*c*	59–60, 63, 70.
Multiliteral systems, historical examples of	36	61.
Normal distribution	17*b*, *c*	27, 28.
Normal frequency	9, 11, 25	11–13, 16–17, 41.
Deviations from	13*b*	19.
Nulls	40, 47*c*	68–69, 99.
New York Tribune, ciphers in	36	61.
Patterns	33*d*	56.
Phraseology of messages	49*a*	101.
Plain component, completion of	20*a*	34.
Plain-text unit	41*c*	70.
Playfair cipher	44, 46	73–83, 84–98.
Modified	46*d*	86.
Polyalphabetic cipher distinguished from monoalphabet	12, 14	18, 22–25.
Polygraphic substitution	41	70–71.
Prefixes in triliteral distribution	27*e*	44.
Prerequisites for cryptographic work	2	2–5.
Probable-word method	33	55–57.
Pseudo-polygraphic systems	41*e*	71.
Punctuation in telegraphic text	10*c*	15.
Quadraliteral cipher	35*d*	60.
Quinqueliteral cipher	35*e*	60.
Random text, number of blanks	14*f*	24.
Relative frequencies	10*b*, *c*, *d*, 11, 14*b*	15, 15, 15, 16–17, 23.
Repetitions	13*g*, 24*b*, *c*, 27	21, 40, 41, 43–46.
In a code message	38*c*	64.
Of consonants	32*c*	53.
Of digraphs and trigraphs	27*f*	46.
Condensed table of	27*i*	46.
Reversed standard alphabet	16, 20*b*	26, 36.
Reversible digraphs indicated on worksheet	26*f*	43.
Russian Morse alphabet	5*b*, 48*e*	8, 101.
Security of monoalphabet using standard alphabets	23	40.
Sequences:		
Known	23	40.
Mixed	21*d*	39.
Unknown	23	40.

	Paragraphs	Pages
Solutions of a subjective nature	3	5.
Specific key	4, 7, 19*a*, 31*b*	7, 10, 31, 52.
Standard alphabets	15*a*, 16, 20*b*, 38*e*	25, 26, 36, 65.
Subjective solutions	3	5.
Substitution:		
Biliteral	41	70–71.
Digraphic	41*c*, 42*a*	70, 72.
Distinguished from transposition	12, 13	18, 18–22.
Polygraphic	41*c*	70.
Multiliteral	41*c*	70.
Trigraphic	41*c*	70.
Triliteral	41	70–71.
Suffixes in trigraphic distribution	27*e*	44.
Suppression of frequency	37, 40*b*, 41*f*	63, 69, 71.
Symbols as cipher elements	13*h*, 48	22, 100–101.
Telegrams, average length of	11*b*	16.
Terminology	1	1.
Text, different types of	10*b*, *c*	15, 15.
Transposition distinguished from substitution	12, 13	18, 18–22.
Trigraphic cipher system	41*c*	70.
Trigraphic frequency table	27 (footnote 2)	44.
Triliteral frequency distribution	27	43–46.
Triliteral substitution	41	70–71.
Type numbers for cryptographic systems	50*f*	104.
Uniliteral frequency distribution	9, 17	11–13, 27–28.
Unknown sequences	23	40.
Variants	37*d*, 40*b*, 49*c*, *d*, *f*	63, 69, 101, 102, 102.
Vowels:		
Average distance apart	32*c*, footnote	53.
Combinations with consonants	28, 29	46–47, 48–49.
Combinations with vowels	29*a*	48.
Distinguished from consonants	28, 32*c*	46–47, 53.
In succession	32*c*	53.
Relative frequency of	10*a*, 13, 19	13, 18–22, 31–33.
Word formulas	33*d*	56.
Word lengths in a cryptogram	26*c*, 32*e*, 33*d*, *f*, *g*	42, 54, 56, 56, 57.
Word patterns	33*d*	56.
Word skeletons	30*b*, 32*e*	49, 54.
Work sheet, preparation of	26	42–43.

ANALYTICAL KEY FOR CRYPTANALYSIS[1]

(Numbers in parentheses refer to paragraph numbers in this text)


CRYPTOGRAMS (4-6, 38b, c, 48)
101—Cipher (41 42)
102—Code (to be taken up later)
201—Polygraphic
202—Monographic
301—Substitution
302—Transposition (to be taken up later)
303—Transposition (to be taken up later)
304—Substitution (12-16)
305—Combined substitution-transposition (to be taken up later) (17d, 22b)
401—Monoalphabetic
402—Polyalphabetic (to be taken up later)
403—Monoalphabetic (35)
404—Polyalphabetic (to be taken up later)
501—Random tabular equivalents (41g)
502—Systematically derived equivalents (41-46)
503—Uniliteral (41)
504—Multiliteral (35-36)
601—Digraphic
602—Trigraphic
603—Etc.
604—Digraphic (42-46)
605—Trigraphic (41g)
606—Etc. (41g)
607—Ciphers composed exclusively of letters (15-24)
608—Ciphers composed of figures, symbols, or intermixtures of letters, figures, and symbols. (48a-b)
609—Single equivalent substitution (35-36)
610—Multiple-equivalent substitution (37-40)
701—Sliding alphabets (41e-f)
702—Checkerboard designs (44-46)
703—Standard cipher alphabet (15-22)
704—Mixed cipher alphabet (25-34)
705—Number ciphers (36b, c)
706—Letter ciphers (35a-e, 36a)
707—Number ciphers (38)
708—Letter ciphers (39-40)
801—Four-square designs (44)
802—Two-square designs (45)
803—One-square designs (Playfair) (46)
804—Direct (19a, 20a)
805—Reversed (19b, 20b)
806—Systematically mixed
807—Random mixed (41c-e)
808—4-Alphabet schemes with numbers from 00-99 (38a-e)
809—Rectangles (39-40)
810—Book method (40e)
901—Keyword mixed (31c-d)
902—Other types (31e)

(Face p. 142)

---

[1] For explanation, see Par. 60.

PROBLEMS

1. In his classic treatise MANUAL FOR THE SOLUTION OF MILITARY CIPHERS Captain Parker Hitt said that there were four essential things necessary for cryptanalytic success. What were they?

2. Five additional elements might be added to the four ingredients that Hitt thought essential to cryptanalytic success. What are they?

3. Normally four fundamental operations or steps are involved in reaching the solution of a cryptogram. What are they?

4. In what language would you expect the cryptographed portion of the following message to be written? Give the reasons for your answer.

From: Pedro Muñetones, Lima, Peru
To: Eduardo Consuelo, Philadelphia

PARA SR. GONZALES. QBDMZIB ZMCPMCPI UIK
YM VCSIAZMOVIC MY BQCIA MYSAQPI DIZMB QC
YIB MCTQYQB QB VZFIADMCDQ FIA BEFEQBDI REQ
YI DIPI REQPQ OICSVPQCOVMY.

5. In the final analysis concerning the solution of cryptograms involving a form of substitution to what simple terms must the cryptograms be reduced in order to reach a solution?

6. In order to reconstruct plaintext is it always necessary to determine the specific key used by the correspondents?

7. The words of sentences in the following examples have been transposed. Rearrange the words to make the probable original plaintext.

(1) STRENGTH ATTACKED HAS IN ENEMY

(2) ARE LIGHT EXTREMELY CASUALTIES

(3) HOURS ATTACK SIX AT OUR HUNDRED BEGAN

(4) FRONT IS ON ALL WESTERN THE QUIET

8. Letters of various words in the sentences below have been transposed. Rearrange the letters to make words that will provide the likely plaintext.

(1) First GRDABIE has now DCAHERE a point NEESV miles East of EIRVR.

(2) Our SOERFC are now DRRAEPPE to launch attack TGINAAS enemy tanks.

(3) Still no word ORFM our intelligence OCUSSRE concerning YEEMN.

9. In the following sentences the words have been transposed and in the case of several words the letters have been transposed. Reconstruct the probable plaintext.

(1) WILL NWAD ATTACK ECOMNCME AT

(2) HAS EEOBCM FIRE YEVHA ENEMY VERY YRTLEAIRL RECENTLY

10. In the following example the letters of each word have been rearranged into the order in which they appear in the normal standard alphabet. Reconstruct the plaintext.

ACCDGINOR OT AAEIILNNNORTT EOPRRST EEMNY ILLW AKS FOR ACEEP

11. In the following examples the plaintext has been broken into five-letter groups and each group of five-letters has been then rearranged in the order in which they appear in the normal standard alphabet. Reconstruct the plaintext.

(1) CMOOT ADIMN EGGNN AEFLR DIRST IIISV NOOST EMOPV HORUY ADEQU AERRT
OSSTU FHIOR ERVXX

(2) ACEHV OOPRR AILLW AILMS EOPRR HOTTT AEHIS ADQRU AERST NOOSS AOPSS
BEILS

12. Using cross-section paper with 1/4 inch squares make a monoliteral frequency distribution of the letters in the following news item:

THE NATIONAL GEOGRAPHIC SOCIETY REPORTS THAT DESPITE ITS NAME
THE NEW RIVER OF WEST VIRGINIA IS ACTUALLY THE OLDEST RIVER
IN NORTH AMERICA MAINTAINING ITS ANCIENT COURSE FOR A HUNDRED
MILLION YEARS.

13. What percentage of the above text are the six vowels, A E I O U Y? How does this compare with the expected frequency of vowels in plaintext?

14. What are the four most frequent consonants in English telegraphic text?

15. What are the five least frequent letters in English telegraphic text?

16. Why are the frequencies of letters in English literary text different from those in telegraphic text?

17. What four facts can be determined from a study of the monoliteral frequency distribution of a cryptogram?

18. Determine the class of cipher system (substitution or transposition) used to encipher the following two messages:

(1) RTOIR MIIST ETUDL PTLVS QHPEI AHLPT UARCE NAEPO ETOHC YRROP SYVAO
WTYRX

(2) VAPER NFRQC ERFFH ERPBA GVAHR FNTNV AFGBH EYVAR FOLGU RRARZ LFGBC

19. Which of the following substitution ciphers are monoalphabetic?

(1) KHNTV OEJGD RXSHV TYDEU USQAJ SFCAZ YXWAN OFWAK XTRIL TBXFS KZOSR
QKHBV JOQOG GVRWB PKOWY RTSBH KSEBW RNYAB QOUOG KVWAV DOADJ OLYLC
CTWKF ATXRI LFRVG FGZGH UOKVR GVEIG JHRXK OFYGC AGKDB YKWOR WOOIV

(2) BVPBY EPBUE NWDCP BVNGJ WNTGC WDWGN XGBVK TVRDV UPCCP EHDCC DGTJV
CDWHH JWHHH WSVHL NVVUP WBCET TGBDW XVNVI LCDBW SVCPC CECTV RDVUC
DGTSV VTECW NXGBI VUWXP NLDZW NYNVJ UVFVH GTVCK

(3) SXSOK TECPS DMCSO MDDSQ DEGFV SMDFE TVWBB FIEDW JTWHS LSCED FERIS
VMHSP MRFGC SQEXS RCWDE XSCNG FVSCS TGDSD FEFMB ZDFER MWCPC MTFIW
BBFMZ SETTM FQMIX FEHSC WTKSX SOKBE PMFWE XDMNP

20. Each of the following messages has been monoalphabetically enciphered. In each case determine whether the cipher alphabet used was a standard alphabet or a mixed alphabet; and if standard, whether direct or reversed.

(1) YLNLC WLULC WXHCM HNPWL WIPWW ILLCL DRTHE EAYBO POERA VEEOP NFHWX
KBYNL XHCWI LCLPY KVWVY LXWBA OVWNB CWHCV LRBVY UHJHE LCNLS

(2) VBEJE FSVWA FJABE IMBTM YBDAS SAWGM FSMNM REGFS MADWM LPGFQ VMJGS
GMFMF SXEBG QXSIW AFPMI SXEJG KSXBE QGNEF SAWLM NDASS EANJS MVVWE
AJEAU RGJEG ITMYL AFLMN VWESE SXENM REGFS XGBST JGKXM YBJJS MVQMM
UWYLP

(3) XCITA AXVTC RTXCS XRPIT HIWPI TCTBN UDGRT HUPRX CVNDJ GGXVW IUAPC
ZLXAA QTGTE APRTS QNUGT HWIGD DEHLW XRWWP KTYJH IPGGX KTSUG DBIGP
XCVRT CITGH HIDEL XAAZT TENDJ PSKXH TSVVV

21. Solve the following cryptogram:

ZWLJS YQDWJ VZNWJ FRRZS NYNTS XYTUJ SJRDN XJCUJ HYJIY TWJXZ RJYMJ
NWTKK JSXNA JXMTW YQDXY TUBJB NQQFQ RTXYX ZWJQD SJJIW JUQFH JRJSY
XXTTS XYTUF IANXJ BMJSF RRZSN YNTSB NQQGJ XJSYC

22. What is the "specific key" used to encipher the message in Problem No. 21?

23. What are the two methods for solving ciphers of the type in Problem No. 21?

24. What are the successive steps used in solving a substitution cipher by completing the plain component sequence where reversed standard alphabets are involved?

25. Solve the following cryptogram using the steps shown in your answer to Problem No. 24:

WSPQN LERLL XELGQ KCAEM ALNER MSWLL WRYDG NEBWQ

26. If a short cryptogram is enciphered by using a cipher alphabet where the plain component is a known mixed sequence, can it be solved by completing the plain component sequence?

27. In attempting to solve an unknown cryptogram that is obviously a substitution cipher what is the first solution technique that the cryptanalyst should try?

28. Solve the following cryptogram:

TGRNA KUPGI CVKXG

29. In attacking a cryptogram that is obviously a monoalphabetic substitution cipher, if solution by completing the plain component sequence is unsuccessful, what type of cipher alphabet may be assumed to have been used?

30. Monoalphabetic substitution ciphers which involve standard cipher alphabets have an extremely low degree of cryptographic security. Why?

31. The following cryptogram has been intercepted. Construct a triliteral frequency distribution of the letters in the cryptogram, showing the preceding and following letter for each letter in the cryptogram. Then indicate by underscoring in the cryptogram all repetitions of three or more letters.

SYSWI LCCLP NPLYM SSHRS PCSQL YICXW SBCOR XTGSA SPSXF SLYIB RSPXT
XPYSG BOAXY TOAWL CXOYR SACLX YXYUC OSYSW IBORS ALCXO YKDOC SBDYB
VXYSD YKDOC SGSGX EEKDX PNILQ FXBSI ODBCO RXYWS LYCXW SNSSR LESAC
COLYI PVLYU SXYCO YSOTS YSWIR AORLU LYQLM AOLQP LBCBB COR

32. With respect to the cryptogram in Problem No. 31, prepare a condensed table which shows the repetitions of trigraphs and digraps that appear more than twice in the cryptogram. Then using the data obtained, solve the cryptogram. Finally, determine the cipher alphabet and the keyword used.

33. What two places in a message lend themselves more readily to attack by the assumption of words than do any other places? Why?

34. What is meant by the "probable word method" of solution?

35. What is meant by the word formula ABCADB? Does the word "people" fit this pattern?

36. For each formula given below indicate a good English word that fits the pattern.

(1) ABBA

(2) ABBACD

(3) ABCA

37. Using the cipher alphabet which pertains to the cryptogram given in Problem 31, by means of it solve the following cryptogram.

NQQNW DNQIH WOOHO GXBPS HJMOJ NMOY

38. Solve the following cryptogram:

GTRTC IXCIT GRTEI TSBTH HPVVT HXCSX RPITT CTBNX HEAPC CXCVP CDUUT
CHXKT CTMIL TTZ

39. Solve the following monoalphabetic substitution cipher.

ZEQEG GSFPK FMENN KQOBN KBAZR SZZSH KEFAZ EDBSP KEAKH OFQOW KHHRO
KFONN OQZNB EGTOB EAKVL YFPBO PLEYB AZEGG EBEWA ZEDYA OGOAA OFMOB
AOBXK QOSZZ LSZZK GOVVV

40. Solve the following: ZHDUH XQGHU KHDYB DWWDF NVWRS.

41. Solve the following cryptogram:

FZELZ HJJOV JVGGD ZFHCB BZGWZ SLFBK HOZYY CXLAZ BJVJK CBZBV WGKBM
JOZAJ CDVHH JOFCL MOXVD KJVGC SBZLJ FVGXC LBJFT YZHKM BVJZY KBTCL
FCFYZ FHVHB VJKCB ELCJZ UZWFV LBELC JZJOZ TQKGG FZAVK BZBFC LJZSC
FCBGT JQZBJ TOCLF HHJCD VYPKH ZQOZB QZXVB DKXNL DYCXL AZBJH

42. Solve the following cryptogram:

RBQBD DVCPI CKKMC MNVEL ILRJP ISIOI BCVRR MCRIB CICRM EEIKM CQMBL
LIQMN ORBAZ MMFEX ICRME EIKMC QMOTD DVNXL NBDVN DXJMV PGTVN RMNOI
CPIQV RMOXB TQVCM YAMQR KNVPT VEMCM DXZIR JPNVZ VELNB DXBTN NIKJR
LEVCF ORBAC BCMRJ MEMOO MCMDX ORIEE JVOIC RMCRI BCOBL VRRVQ FICKU
ETLLB SMNEB BFICK NISMN

43. The following message is believed to be enciphered in the same general system, but with a different specific key, as the message in Problem No. 42:

SMSPA WHNFF NRUWM YWMZL IRS

44. The following cryptogram is an example of a biliteral, monoalphabetic substitution system. The cipher alphabet is said to be bipartite. Solve the cryptogram and determine the keyword by which the bipartite alphabet is derived.

BBNLN NLNOB OBBDN LDNLB LNNLB DLNNO OLNDB OLOOB BDNON OLNOO NNBDO
ONOBD NNLOB BOOLN NNDLB BNONO BDDBB DNLBB OBLOO NNDNL NOBBL LONNO
NNONO DONND NLBOB BNOND BBOBN NLNBD NOLOB BDBBD NOONL BBBNL BLBDB
DOOON OOBDO LBBOO DLONN DOOBN BDBNN ONNON OD

45. From a cryptographic viewpoint what is a major disadvantage of a multiliteral substitution system?

46. Solve the following cryptogram:

ROBOW SNHBO OSWEN SWURS WHBSW UWERE BSWEB ONHBU NHWSW EWERS BEWEO
HWSWE BSWHW EBHRS OSNOR SOUOO OSOSR SRUWS OONUW ERSRS BEBHB SNHBE
RURSR UREBS NHWUO EOOWS WSOOB EBUWS WEBSW HRSNO BORON HBOWE OOBSB
EBSRE BHBSN HBERU RSRUO HBOWS BEBSW ERHRS RSBEB ORORO BSOEW HOSOO
WSOHR SRURU NHRSW EBSOH RSBON ONSBO WUWEO OOSOS RSWUN SREOO WURSB
HOHOO ROOHO HBOWS WHWUR SNORS BEWER SRUBO OOWUR OWUBO REWER EWUBS
OEOSB OBERU OOBEB UWSWE BSWH

47. Solve the following cryptogram:

28274 94927 38174 63019 10272 94740 27264 91026 48571 73838 28103
94656 10294 95710 38561 03917 38104 81027 48494 62017 49481 82610
48103 04938 46162 74917 46304 84946 18394 65610 57173 83828 10392
72910 57174 91917 30495 71030 49592 04640 26194 64026 48484 94618
26104 74010 48494 91927 49491 91748 19102 72947 40272 64910 26482
81030 46491 72017 10292 72049 10263 94656 10484 94618

48. What is the value of having variants in a monoalphabetic substitution cipher system?

49. Explain the difference between a simple monoalphabetic substitution cipher system and a monoalphabetic substitution system with variants.

50. Solve the following cryptogram:

21 30 31 70 42 30 79 68 70 74 22 52 26 95 96 44 18 49 54 60 21
57 09 07 70 17 23 78 70 43 58 05 72 96 08 80 38 94 70 83 79 13
17 11 70 48 68 09 84 22 38 24 16 67 70 83 46 09 73 74 28 15 96
43 13 53 85 33 34 22 08 13 25 74 84 00 80 17 58 59 54 55 21 96
82 60 09 43 85 23 12 66 44 53 18 10 00 42 16 29 09 07 74 84 00
54 79 93 09 52 05 69 96 94 39 17 68 70 42 38 34 79 32 44 33 13
22 37 66 44 59 96 57 86 53 59 74 51 05 71 23 70 57 85 73 96 26
57 83 74 61 92 51 66 23 03 18 45 83 99 30 05 29 20 45 66 57 44
30 21 22 18 97 54 86 57 77 13 05 00 84 39 17 80 97 10 34 07 96
21

Probable words: REFERENCE, DIVISION, HEADQUARTERS, OFFICER.

51. Solve the following cryptogram:

26211 55185 25108 81813 58698 11434 63651 18756 45312 88691 64335
72326 27613 96325 11588 63082 15215 53230 56866 16025 83868 18426
75318 23456 17102 78541 33238 45210 58281 38819 60281 02686 30158
06039 30635 85386 67335 22251 88312 36348 17402 86857 11388 35681
24951 32282 91217 05756 61108 53261 70542 68123 86871 12819 26171
35821 27686 15756 51165 16388 33283 05626 60355 02990 31877 16038
42132 33183 81218 81153 86

Probable words: COMMANDING OFFICER.

52. The following two cryptograms have been intercepted. Since the enemy has been using "stereotype" message beginnings, it is believed that both cryptograms begin the same. Probable words that can be expected in the text include: ENEMY, DIVISION, STOP, REPORT, ATTACK. Solve both cryptograms.

No. 1:

BH QF FM XF PD GL VP ZX TW GF XJ LM SX MH CM GS TF SV PX HT GS LX VN
BH MN GS KF SD NB SD QS JX WV HV QS XN XL VS QW XS LQ HV VS QJ FV DL
RC LV ST VH FD SR SD CB HK XS QZ JD SG HD GC KP VF PQ FQ VS FD ML ZD
HB PK VS QL SX PK BS XL XF

No. 2:

HB FX MS QS DP LD VL QJ MN GS ZQ LM XS TH CM FD ST BF XL MC GS XP VN
BH MW FG SK GF BW SD XF QZ WV HV FX WQ VS JQ VS PD HK VL CV HB LB MS
ND DS BN VS RW XN BF ST HV VS MN GS JQ CV JD FV ZX SB NR HB HV VJ BF
DH RF VS GJ XS KZ QL VN

52. The enemy has suddenly started to use a new cryptographic system. This is one of the cryptograms in the new system. Solve the cryptogram.

SDLUC EUIYA KTNIC UDISR FYTYA NLNYR AFFUD IZSKM ANYBS DLRLV ICTGO
PHILO BHDSC CMCSU FLSDB NAOTR YGZMK ESQTL OUGCU IYAKT NIAEA OCKPM
EOMLF LUTRP MDCOG LOKHS PIOWM UFRNL LUIDV MSMKF URUMB LAORY KHFUT
LNUED PILDA NLUBF GLPIS X

Probable message ending: GENERAL JONES.

53. What is the difference between a polyliteral substitution cipher system and a polygraphic substitution cipher system?

54. The Playfair cipher is perhaps the most well-known digraphic substitution cipher system. What characteristics in the ciphertext make the system recognizable?

55. Solve the following cryptogram:

CLSQN FRPPR GKPOX HCBSF TUKSO HVFKD OZNLC VBMCP LZIKE VPONL RMPTS
FUAGE MTFSU MOWMB ITCZC LDBLJZ SFTPF ZNLGU GBDKU ARSMF SQRQU MPORQ
TPUVU GFHCV

Probable words: BATTALION, STOP, HEADQUARTERS.

56. In solving cryptograms involving symbols what is the usual preliminary procedure that the cryptanalyst follows?

57. Solve the following cryptogram:

δ ε α θ ε τ θ α ν σ μ ι τ τ ι ν λ ο ν α ɩ θ ε γ ξ ε
ν ψ υ ο ɩ τ δ ο σ ι χ ζ ε θ ο ɩ ι × ε κ ι ω ο ψ υ ψ
ω ε σ σ τ ο ρ ρ ω ε α σ ε α φ × ι σ ε ι ɩ υ ο ξ ψ α
ν θ ε ψ ε ι × ε ο ξ θ σ ι λ ν α ω σ σ τ ο ρ ε ν ε μ
υ ι σ σ ξ θ ε ω υ ι ν τ ε θ ψ ε ρ τ ι ν λ ο ξ θ μ ε
σ σ α λ ε σ σ ο ξ σ ε ψ α θ ε ι ν υ ο ξ θ τ θ α ν σ
μ ι σ σ ι ο ν σ σ τ ο ρ σ ι λ ν ε φ λ ε ν ε θ α ω δ
ι ω ω ι α μ σ

58. Solve the following cryptogram:

AXXAK GOASD VVREA MRKOV QDUVR VLUSC TNSXQ TBTST WRSXW ITRTX TAEKA
SMAEX TVSAN VETHO XSXWI JVARX TKTIA XVXOA XXOVV RVLUJ TEEAQ QXJWL
WNVQT BTSTW RSXWX OVAXX AKGTR KEMQT RHANX TEEVN USMII WNXCN WLANL
UMRTX SSXWI JVANV KWRXT RMTRH XWLAG VAVNT AEWDS VNBAX TWRST RSITX
VWCVR VLUAT NAKXT BTXUS XWIJT EEGVV IUWMA QBTSV Q

59. What is the analytical key for cryptanalysis?

C-1	MANUAL FOR THE SOLUTION OF MILITARY CIPHERS, Parker Hitt
c-3	ELEMENTS OF CRYPTANALYSIS, William F. Friedman
c-4	STATISTICAL METHOD&N CRYPTANALYSIS, Solomon Kullback
C-5	CRYPTOGRAPHY AND CRYPTANALYSIS ARTICLES, Vol. 1, Friedman
C-6	CRYPTOGRAPHY AND CRYPTANALYSIS ARTICLES, Vol. 2, Friedman
c-9	WAR SECRETS IN THE ETHER, Vol. 1, Wilhelm F. Flicke
c-10	WAR SECRETS IN THE ETHER, Vol. 2, Wilhelm F. Flicke
c-17	CRYPTANALYSIS OF THE HAGELIN CRYPTOGRAPH, Wayne G. Barker
C-18	THE CONTRIBUTIONS OF THE CRYPTOGRAPHIC BUREAUS IN THE WORLD WAR [World War I], Yves Gylden
c-20.	HISTORY OF CODES AND CIPHERS IN THE U.S. PRIOR To WORLD WAR I, ed. Barker
c-21	HISTORY OF CODES AND CIPHERS IN THE U.S. DURING WORLD WAR I, ed. Barker
c-22	HISTORY OF CODES AND &HERS IN THE U.S. DURING THE PERIOD BETWEEN THE WORLD WARS, PART I. 1919-1929, ed. Barker
c-30	MILITARY CRYPTANALYSIS, PART I, William F. Friedman
c-33	COURSE IN CRYPTANALYSIS, Volume 1, British War Office
c-34	COURSE IN CRYPTANALYSIS, Volume 2, British War Office
c-35	THE ORIGIN AND DEVELOPMENT OF THE NATIONAL SECURITY AGENCY, Brownell
c-39	CRYPTANALYSIS OF SHIFT-REGISTER GENERATED STREAM CIPHER SYSTEMS, Wayne G. Barker
c-40	MILITARY CRYPTANALYSIS, PART II, William F. Friedman
c-41	ELEMENTARY COURSE IN'PROBABILITY FOR THE CRYPTANALYST, Gleason
C-42	MILITARY CRYPTANALYTICS, PART I, VOL. 1, L.D. Callimahos
c-43	MILITARY CRYPTANALYTICS, PART I, VOL. 2, L.D. Callimahos
c-44	MILITARY CRYPTANALYTICS, PARTII, VOL. 1, L.D. Callimahos
c-45	MILITARY CRYPTANALYTICS, PART II, VOL. 2, L.D. Callimahos
C-46	PATTERN WORDS: THREE-LETTERS TO EIGHT-LETTERS IN LENGTH, Carlisle
C-48	PATTERN WORDS: NINE-LETTERS IN LENGTH, Sheila Carliie
c-49,	THE INDEX OF COINCIDENCE AND ITS APPLICATIONS IN CRYPTANALYSIS, William F. Friedman
c-50	CRYPTOGRAPHIC SIGNIFICANCE, OF THE KNAPSACK PROBLEM, Luke J. O'Connor and Jennifer Seberry
C-52	THE AMERICAN BLACK CHAMBER, Herbert 0. Yardley
c-53	TRAFFIC ANALYSIS AND THE ZENDIAN PROBLEM, L.D. Callimahos
c-54	HISTORY OF CODES AND CIPHERS IN THE U.S. DURING THE PERIOD BETWEEN THE WORLD WARS, PART II, 1930-1939, ed. Barker
C-55	INTRODUCTION To THE ANALYSIS OF THE DATA ENCRYPTION STANDARD (DES), Wayne G. Barker
C-56	ELEMENTARY CRYPTOGRAPHY AND CRYPTANALYSIS, Donald D. Mill&in
c-57	SECRET CIPHERS OF THE 1876 PRESIDENTIAL ELECTION, D. Beaird Glover
C-58	SOLVING CIPHER PROBLEMS, Frank W. Lewis
C-59	CRYPTANALYSIS OF THE SINGLE COLUMNAR TRANSPOSITION CIPHER, W. Barker
C-60	MILITARY CRYPTANALYSIS, PART III, William F. Friedman
C-61	MILITARY CRYPTANALYSIS, PART IV, William F. Friedman
C-62	PATTERN WORDS: TEN-LEES AND ELEVEN-LETTERS IN LENGTH, Wallace
C-63	PATTERN WORDS: TWELVE-LETTERS AND GREATER IN LENGTH, Wallace

AEGEAN PARK PRESS, P.O. BOX 2837, LAGUNA HILLS, CALIFORNIA 92654

A **CRYPTOGRAPHIC SERIES**

MILITARY CRYPTANALYSIS
PART II

With Added PROBLEMS and COMPUTER PROGRAMS

Pages

Introductory remarks 1-3
Cipher alphabets for polyalphabetic substitution . . 4-9
Theory of solution of repeating-key systems . . 10-16
Repeating-key systems with standard cipher alphabets . . 17-23
Repeating-key systems with mixed cipher alphabets, I . 24-48
Repeating-key systems with mixed cipher alphabets, IT . . 49-51
Theory of indirect symmetry of position in secondary alphabets 52-59
Application of principles of indirect symmetry of position. . 60-77
Repeating-key systems with mixed cipher alphabets, III 78-83
Repeating-key systems with mixed cipher alphabets, IV .. 84-95
Appendix 1 .. 96-107
Appendix 2 . 108-118
Appendix 3 . 119-126
Index . 127-128
PROBLEMS . 132-145
COMPUTER PROGRAMS . 146-158

by

William F. Friedman

From Aegean Park Press

MILITARY CRYPTANALYSIS II 1
With Added PROBLEMS and COMPUTER 1
FOREWORD 4
POLYALPHABETIC SUBSTITUTION SYSTEMS 5
SECTION I.. 6
INTRODUCTORY REMARKS 6
SECTXON II 9
CIPHER ALPHABETS FOB POLYALPHABETIC..................... 9
SECTION III.. 15
THEORY OF SOLUTION OF REPEATING-KEY SYSTEMS 15
SECTION IV 22
REPEATING-KEY SYSTEMS WITH STANDARD CIPHER...... 22
SECTION V 29
REPEATING-KEY SYSTEMS WITH MIXED CIPHER 29
SECTION VI 54
REPEATING-KEY SYSTEMS WITH MIXED CIPHER 54
SECTION VII 57
THEORY OF INDIRECT SYMMETRY OF POSITION IN.......... 57
SECTION VIII 65
APPLICATION OF PRINCIPLES OF INDIRECT...................... 65
SECTION IX 83
REPEATING-KEY SYSTEMS WITH MIXED CIPHER 83
SECTION X 89
BEPEATING-KEY SYSTEMS WITH MIXED CIPEEB............... 89
Analytical Key for Military Cryptanalysis, Part II * 100
APPENDIX 1 101
THE 12 TYPES oF CIPHER SQUARES.................................. 101
APPENDIX 2 113
ELEMENTARY STATISTICAL THEOEY APPLICABLE TO 113
APPENDIX 3 126
A GRAPHICAL METHOD OF RECONSTRUCTING 126
INDEX.. 137
PROBLEMS.. 139
COMPUTER PROGRAMS 153
INDEX OF PROGRAMS.. 154
Vigenere Encipherment 155
TRUE BEAUFORT ENCIPHERMENT..................................... 157
VARIANT BEAUFORT ENCIPHERMENT 159
DETERMINING THE PERIOD OF A PERIODIC CIPHER 161
VIGENERE ENCIPHERMENT USING MIXED ALPHABETS.... 163

MILITARY CRYPTANALYSIS

Part II

SIMPLER VARIETIES
OF POLYALPHABETIC SUBSTITUTION SYSTEMS

By

WILLIAM F. FRIEDMAN

© 1984 Aegean Park Press

ISBN: 0-89412-064-6

AEGEAN PARK PRESS
P. Q. Bsx 2837
Laguna Hills, California 92654
(714)586-8811

Manufactured in the United States of America

FOREWORD

We are proud to add this book, MILITARY CRYPTANALYSTS, PART II, recently declassified by the U.S. Government, to our Cryptographic Series. As in the case of MILITARY CRYPTANALYSIS, PART I, we have added a large number of problems to the book. These problems, largely keyed to the order that the material is presented in the text, not only will provide the student with many hours of enjoyment, but at the same time will act as the ultimate teaching aid.

In keeping with what might be termed modern cryptologic advances, we have also added to the book some computer programs. There is no doubt that the computer has greatly affected modern cryptology, and today cryptographic and cryptanalytic "tasks" which at one time took hours and even days to accomplish can now be done in seconds, if not microseconds. The added computer programs, found at the end of the book, following the problems, are only representative of the many programs that can be used with the large class of cipher systems discussed in this book. The student should set his sights on modifying, improving, and developing other programs which will assist him in his solution efforts.

Comments concerning this book, or any book in our Cryptographic Series, are always greatly received.

September 1984 AEGEAN PARK PRESS

MILITARY CRYPTANALYSIS. PART II. SIMPLER VARIETIES OF POLYALPHABETIC SUBSTITUTION SYSTEMS

Section		Paragraphs	Pages
I.	Introductory remarks	1 4	1–3
XI.	Cipher alphabets for polyalphabetic substitution	5-7	4–9
III.	Theory of solution of repeating-key systems	3-12	10–16
IV.	Repeating-key systems with standard cipher alphabets	13-16	17-23
V.	Repeating-key systems with mixed cipher alphabets, I	16-26	24–48
VI.	Repeating-key systems with mixed cipher alphabets, II	27-30	49–51
VII.	Theory of indirect symmetry of position in secondary alphabets	31	52-69
VIII.	Application of principles of indirect symmetry of position	32-36	60–77
IX.	Repeating-key systems with mixed cipher alphabets, III	3740	78–83
X.	Repeating-key systems with mixed cipher alphabets, IV	4146	84–95
	Appendix 1		96-107
	Appendix 2		108–118
	Appendix 3		119–126
	Index		127-128

SECTION I

INTRODUCTORY REMARKS

Paragraph

The essential difference between monoalphabetic and polyalphabetic substitution 1

Primary classification of polyalphabetic systems 2

Primary classification of periodic systems 3

Sequence of study of polyalphabetic systems 4

1. **The essential difference between monoalphabetic and polyalphabetic substitution.**—*a.* In the substitution methods thus far discussed it has been pointed out that their basic feature is that of monoalphabeticity. From the cryptanalytic standpoint, neither the nature of the cipher symbols, nor their method of production is an essential feature, although these may be differentiating characteristics from the cryptographic standpoint. It is true that in those cases designated as monoalphabetic substitution with variants or multiple equivalents, there is a departure, more or less considerable, from strict monoalphabeticity. In some of those cases, indeed, there may be available two or more wholly independent sets of equivalents, which, moreover, may even be arranged in the form of completely separate alphabets. Thus, while a loose terminology might permit one to designate such systems as polyalphabetic, it is better to reserve this nomenclature for those cases wherein polyalphabeticity is the essence of the method, specifically introduced with the purpose of imparting a *positional* variation in the substitutive equivalents for plain-text letters, in accordance with some rule directly or indirectly connected with the absolute *positions* the plain-text letters occupy in the message. This point calls for amplification.

b. In monoalphabetic substitution with variants the object of having different or multiple equivalents is to suppress, so far as possible by simple methods, the characteristic frequencies of the letters occurring in plain text. As has been noted, it is by means of these characteristic frequencies that the cipher equivalents can usually be identified. In these systems the varying equivalents for plain-text letters are subject to the free choice and caprice of the enciphering clerk; if he is careful and conscientious in the work, he will really make use of all the different equivalents afforded by the system; but if he is slip-shod and hurried in his work, he will use the same equivalents repeatedly rather than take pains and time to refer to the charts, tables, or diagrams to find the variants. Moreover, and this is a crucial point, even if the individual enciphering clerks are extremely careful, when many of them employ the same system it is entirely impossible to insure a complete diversity in the encipherments produced by two or more clerks working at different message centers. The result is inevitably to produce plenty of repetitions in the texts emanating from several stations, and when texts such as these are all available for study they are open to solution, by a comparison of their similarities and differences.

c. In true polyalphabetic systems, on the other hand, there is established a rather definite procedure which automatically determines the shifts or changes in equivalents or in the manner in which they are introduced, so that these changes are beyond the momentary whim or choice of the enciphering clerk. When the method of shifting or changing the equivalents is scientifically sound and sufficiently complex, the research necessary to establish the values of the cipher characters is much more prolonged and difficult than is the case even in complicated monoalphabetic substitution with variants, as will later be seen. These are the objects of true polyalphabetic substitution systems. The number of such systems is quite large, and it will be possible to describe in detail the cryptanalysis of only a few of the more common or typical examples of methods encountered in practical military communications.

d. The three methods, (1) single-equivalent monoalphabetic substitution, (2) monoalphabetic substitution with variants, and (3) true polyalphabetic substitution, show the following relationships as regards the equivalency between plain-text and cipher-text units:

A. In method (1), there is a set of 26 symbols; a plain-text letter is always represented by one and only one of these symbols; conversely, a symbol always represents the same plain-text letter. The equivalence between the plain-text and the cipher letters is constant in both encipherment and decipherment.

B. In method (2), there is a set of *n* symbols, where *n* may be any number greater than 26 and often is a multiple of that number; a plain-text letter may be represented by 1, 2, 3, . . . different symbols; conversely, a symbol always represents the same plain-text letter, the same as is the case in method (1). The equivalence between the plain-text and the cipher letters is variable in encipherment but constant in decipherment.[1]

C. In method (3) there is, as in the first method, a set of 26 symbols; a plain-text letter may be represented by 1, 2, 3, . . . 26 different symbols; conversely, a symbol may represent 1, 2, 3, . . . 26 different plain text letters, depending upon the system and the specific key. The equivalence between the plain-text and the cipher letters is variable in both encipherment and decipherment.

2. **Primary classification of polyalphabetic systems.**—*a.* A primary classification of polyalphabetic systems into two rather distinct types may be made: (1) periodic systems and (2) aperiodic systems. When the enciphering process involves a cryptographic treatment which is repetitive in character, and which results in the production of *cyclic phenomena* in the cryptographic text, the system is termed *periodic*. When the enciphering process is not of the type described in the foregoing general terms, the system is termed *aperiodic*. The substitution in both cases involves the use of two or more cipher alphabets.

b. The cyclic phenomena inherent in a periodic system may be exhibited externally, in which case they are said to be *patent*, or they may not be exhibited externally, and must be uncovered by a preliminary step in the analysis, in which case they are said to be *latent*. The periodicity may be quite definite in nature, and therefore determinable with mathematical exactitude allowing for no variability, in which case the periodicity is said to be *fixed*. In other instances the periodicity is more or less flexible in character and even though it may be deter-

---

[1] There is a monoalphabetic method in which the inverse result obtains, the correspondence being constant in encipherment but variable in decipherment; this is a method not found in the usual books on cryptography but in an essay on that subject by Edgar Allan Poe, entitled, in some editions of his works, *A few words on secret writing* and, in other editions, *Cryptography*. The method is to draw up an enciphering alphabet such as the following (using Poe's example):

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S	U	A	V	I	T	E	R	I	N	M	O	D	O	F	O	R	T	I	T	E	R	I	N	R	E

In such an alphabet, because of repetitions in the cipher component, the plain-text equivalents are subject to a considerable degree of variability, as will be seen in the deciphering alphabet:

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	C			M	G	O			E				K	J	L			H	A	F	B	D				
					U				I					X	N			Q		R						
					Z				S						P			V		T						
									W									Y								

This type of variability gives rise to ambiguities in decipherment. A cipher group such as TIE, would yield such plain-text sequences as REG, FIG, TEU, REU, etc., which could be read only by *context*. No system of such a character would be practical for serious usage. For a further discussion of this type of cipher alphabet see Friedman, William F., *Edgar Allan Poe, Cryptographer*, Signal Corps Bulletins Nos. 97 (July–Sept.) and 98 (Oct.–Dec.), 1937.

minable mathematically, allowance must be made for a degree of variability subject to limits controlled by the specific system under investigation. The periodicity is in this case said to be *flexible*, or *variable within limits*.

3. Primary classification of periodic systems.—*a*. Periodic polyalphabetic substitution systems may primarily be classified into two kinds:

(1) Those in which only a few of a whole set of cipher alphabets are used in enciphering individual messages, these alphabets being employed repeatedly in a fixed sequence throughout each message. Because it is usual to employ a secret word, phrase, or number as a key to determine the number, identity, and sequence with which the cipher alphabets are employed, and this key is used over and over again in encipherment, this method is often called the *repeating-key system*, or the *repeating-alphabet system*. It is also sometimes referred to as the *multiple-alphabet system* because if the keying of the entire message be considered as a whole it is composed of multiples of a short key used repetitively.[3] In this text the designation "repeating-key system" will be used.

(2) Those in which all the cipher alphabets comprising the complete set for the system are employed one after the other successively in the encipherment of a message, and when the last alphabet of the series has been used, the encipherer begins over again with the first alphabet. This is commonly referred to as a *progressive-alphabet system* because the cipher alphabets are used in progression.

4. Sequence of study of polyalphabetic systems.—*a*. In the studies to be followed in connection with polyalphabetic systems, the order in which the work will proceed conforms very closely to the classifications made in paragraphs 2 and 3. Periodic polyalphabetic substitution ciphers will come first, because they are, as a rule, the simpler and because a thorough understanding of the principles of their analysis is prerequisite to a comprehension of how aperiodic systems are solved. But in the final analysis the solution of examples of both types rests upon the conversion or reduction of polyalphabeticity into monoalphabeticity. If this is possible, solution can always be achieved, granted there are sufficient data in the final monoalphabetic distributions to permit of solution by recourse to the ordinary principles of frequency.

b. First in the order of study of periodic systems will come the analysis of repeating-key systems. Some of the more simple varieties will be discussed in detail, with examples. Subsequently, ciphers of the progressive type will be discussed. There will then follow a more or less detailed treatment of aperiodic systems.

---

[3] French terminology calls this the "double-key method", but there is no logic in such nomenclature.

SECTION II

CIPHER ALPHABETS FOR POLYALPHABETIC SUBSTITUTION

Paragraph
Classification of cipher alphabets upon the basis of their derivation........ 5
Primary components and secondary alphabets........ 6
Primary components, cipher disks, and square tables........ 7

5. Classification of cipher alphabets upon the basis of their derivation.—*a*. The substitution processes in polyalphabetic methods involve the use of a plurality of cipher alphabets. The latter may be derived by various schemes, the exact nature of which determines the principal characteristics of the cipher alphabets and plays a very important role in the preparation and solution of polyalphabetic cryptograms. For these reasons it is advisable, before proceeding to a discussion of the principles and methods of analysis, to point out these various types of cipher alphabets, show how they are produced, and how the method of their production or derivation may be made to yield important clues and short-cuts in analysis.

b. A primary classification of cipher alphabets for polyalphabetic substitution may be made into the two following types:

(1) Independent or unrelated cipher alphabets.

(2) Derived or interrelated cipher alphabets.

c. Independent cipher alphabets may be disposed of in a very few words. They are merely separate and distinct alphabets showing no relationship to one another in any way. They may be compiled by the various methods discussed in Section IX of *Elementary Military Cryptography*. The solution of cryptograms written by means of such alphabets is rendered more difficult by reason of the absence of any relationship between the equivalents of one cipher alphabet and those of any of the other alphabets of the same cryptogram. On the other hand, from the point of view of practicability in their production and their handling in cryptographing and decryptographing, they present some difficulties which make them less favored by cryptographers than cipher alphabets of the second type.

d. Derived or interrelated alphabets, as their name indicates, are most commonly produced by the *interaction of* two primary components, which when juxtaposed at the various points of coincidence can be made to yield *secondary alphabets*.[1]

6. Primary components and secondary alphabets.—Two basic, slidable sequences or components of n characters each will yield n secondary alphabets. The components may be classified according to various schemes. For cryptanalytic purposes the following classification will be found useful:

CASE A. The primary components are both normal sequences.

(1) The sequences proceed in the same direction. (The secondary alphabets are direct standard alphabets.) (Pars. 13–15.)

(2) The sequences proceed in opposite directions. (The secondary alphabets are reversed standard alphabets; they are also reciprocal cipher alphabets.) (Par. 13*i*, 14*g*.)

CASE B. The primary components are not both normal sequences.

(1) The plain component is normal, the cipher component is a mixed sequence. (The secondary alphabets are mixed alphabets.) (Par. 16–25.)

---

[1] See Sec. VIII and IX, *Elementary Military Cryptography*.

(2) The plain component is a mixed sequence, the cipher component is normal. (The secondary alphabets are mixed alphabets.) (Par. 26.)

(3) Both components are mixed sequences.

(a) Components are identical mixed sequences.

I. Sequences proceed in the same direction. (The secondary alphabets are mixed alphabets.) (Par. 28.)

II. Sequences proceed in opposite directions. (The secondary alphabets are reciprocal mixed alphabets.) (Par. 38.)

(b) Components are different mixed sequences. (The secondary alphabets are mixed alphabets.) (Par. 39.)

7. Primary components, cipher disks, and square tables.—*a.* In preceding texts it has been shown that the equivalents obtainable from the use of quadricular or square tables may be duplicated by the use of revolving cipher disks or of sliding primary components. It was also stated that there are various ways of employing such tables, disks, and sliding components. Cryptographically the results may be quite diverse from different methods of using such paraphernalia, since the specific equivalents obtained from one method may be altogether different from those obtained from another method. But from the cryptanalytic point of view the diversity referred to is of little significance; only in one or two cases does the specific method of employing these cryptographic instrumentalities have an important bearing upon the procedure in cryptanalysis. However, it is advisable that the student learn something about these different methods before proceeding with further work.

b. There are, not *two,* but *four* letters involved in every case of finding equivalents by means of sliding primary components; furthermore, the determination of an equivalent for a given plain-text letter is representable by *two* equations involving *four* elements, usually letters. Three of these letters are by this time well-known to and understood by the student, viz, θ_k, θ_p, and θ_c. The fourth element or letter has been passed over without much comment, but cryptographically it is just as important a factor as the other three. Its function may best be indicated by noting what happens when two primary components are juxtaposed, for the purpose of finding equivalents. Suppose these components are the following sequences:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Now suppose one is merely asked to find the equivalent of P_p when the key letter is K. Without further specification, the cipher equivalent cannot be stated; for it is necessary to know not only which K will be used as the key letter, the one in the component labeled (1) or the one in the component labeled (2), but also what letter the K_k will be set against, in order to juxtapose the two components. Most of the time, in preceding texts, these two factors have been tacitly assumed to be fixed and well understood: the K_k is sought in the mixed, or cipher component, and this K is set against A in the normal, or plain component. Thus:

```
                               Plain      Index
                                 ↓          ↓
(1) Plain........ ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
(2) Cipher.......         FBPYRCQZIGSEHTDJUMKVALWNOX
                                 ↑          ↑
                              Cipher       Key
```

With this setting $P_p = Z_c$.

c. The letter A in this case may be termed the *index letter*, symbolized A_i. The index letter constitutes the fourth element involved in the two equations applicable to the finding of equivalents by sliding components. The four elements are therefore these:

(1) The key letter, θ_k
(2) The index letter, θ_i
(3) The plain-text letter, θ_p
(4) The cipher letter, θ_c

The index letter is commonly the initial letter of the component; but this, too, is only a convention. It *might be any letter* of the sequence constituting the component, as agreed upon by the correspondents. *However, in the subsequent discussion it will be assumed that the index letter is the initial letter of the component in which it is located, unless otherwise stated.*

d. In the foregoing case the enciphering equations are as follows:

$$\text{(I)}\ K_k=A_i;\ P_p=Z_c$$

But there is nothing about the use of sliding components which excludes other methods of finding equivalents than that shown above. For instance, despite the labeling of the two components as shown above, there is nothing to prevent one from seeking the plain-text letter in the component labeled (2), that is, the cipher component, and taking as its cipher equivalent the letter opposite it in the other component labeled (1). Thus:

```
           Cipher           Index
              ↓               ↓
(1) ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
(2)         FBPYRCQZIGSEHTDJUMKVALWNOX
              ↑               ↑
            Plain            Key
```

Thus:

$$\text{(II)}\ K_k=A_i;\ P_p=K_c$$

e. Since equations (I) and (II) yield different resultants, even with the same index, key, and plain-text letters, it is obvious that an accurate formula to cover a specific pair of enciphering equations must include data showing in what component each of the four letters comprising the equations is located. Thus, equations (I) and (II) should read:

(I) K_k in component (2)=A_i in component (1); P_p in component (1)=Z_c in component (2).
(II) K_k in component (2)=A_i in component (1); P_p in component (2)= K_c in component (1).

For the sake of brevity, the following notation will be used:

$$(1)\ K_{k/2}=A_{i/1};\ P_{p/1}=Z_{c/2}$$

$$(2)\ K_{k/2}=A_{i/1};\ P_{p/2}=K_{c/1}$$

f. Employing two sliding components and the four letters entering into an enciphering equation, there are, in all, twelve different resultants possible for the same set of components and the same set of four basic elements. These twelve differences in resultants arise from a set of twelve different enciphering conditions, as set forth below (the notation adopted in subparagraph *e* is used):

(1) $\theta_{k/2}=\theta_{i/1};\ \theta_{p/1}=\theta_{c/2}$
(2) $\theta_{k/2}=\theta_{i/1};\ \theta_{p/2}=\theta_{c/1}$
(3) $\theta_{k/1}=\theta_{i/2};\ \theta_{p/1}=\theta_{c/2}$
(4) $\theta_{k/1}=\theta_{i/2};\ \theta_{p/2}=\theta_{c/1}$
(5) $\theta_{k/2}=\theta_{p/1};\ \theta_{i/1}=\theta_{c/2}$
(6) $\theta_{k/2}=\theta_{c/1};\ \theta_{i/1}=\theta_{p/2}$
(7) $\theta_{k/2}=\theta_{p/1};\ \theta_{i/2}=\theta_{c/1}$
(8) $\theta_{k/2}=\theta_{c/1};\ \theta_{i/2}=\theta_{p/1}$
(9) $\theta_{k/1}=\theta_{p/2};\ \theta_{i/1}=\theta_{c/2}$
(10) $\theta_{k/1}=\theta_{c/2};\ \theta_{i/1}=\theta_{p/2}$
(11) $\theta_{k/1}=\theta_{p/2};\ \theta_{i/2}=\theta_{c/1}$
(12) $\theta_{k/1}=\theta_{c/2};\ \theta_{i/2}=\theta_{p/1}$

g. The twelve resultants obtainable from juxtaposing sliding components as indicated under the preceding subparagraph may also be obtained either from one square table, in which case twelve different methods of finding equivalents must be applied, or from twelve different square tables, in which case one standard method of finding equivalents will serve all purposes.

h. If but one table such as that shown below as Table I–A is employed, the various methods of finding equivalents are difficult to keep in mind.

TABLE I–A

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X
B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F
P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B
Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P
R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y
C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R
Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C
Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q
I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z
G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I
S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G
E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S
H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E
T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H
D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T
J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D
U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J
M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U
K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M
V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K
A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V
L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A
W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L
N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W
O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N
X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O

For example:

(1) For enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{c/2}$:

Locate Θ_p in top sequence; locate Θ_k in first column;

Θ_c is letter within the square at intersection of the two lines thus determined.

Thus:

$$K_{k/2}=A_{1/1};\ P_{p/1}=Z_{c/2}$$

(2) For enciphering equations $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/2}=\theta_{c/1}$:

Locate θ_k in first column; follow line to right to θ_p; proceed up this column; θ_c is letter at top.

Thus:

$$K_{k/2}=A_{i/1};\ P_{p/2}=K_{c/1}$$

(3) For enciphering equations $\theta_{k/1}=\theta_{i/2}$; $\theta_{p/1}=\theta_{c/2}$:

Locate θ_k in top sequence and proceed down column to θ_i;

Locate θ_p in top sequence; θ_c is letter at other corner of rectangle thus formed.

Thus:

$$K_{k/1}=A_{i/2};\ P_{p/1}=X_{c/2}$$

Only three different methods have been shown and the student no doubt already has encountered difficulty in keeping them segregated in his mind. It would obviously be very confusing to try to remember all twelve methods. But if one standard or fixed method of finding equivalents is followed with several different tables, then this difficulty disappears. Suppose that the following method is adopted: Arrange the square so that the plain-text letter may be sought in a separate sequence, arranged alphabetically, above the square and so that the key letter may be sought in a separate sequence, also arranged alphabetically, to the left of the square; look for the plain-text letter in the top row; locate the key letter in the 1st column to the left; find the letter standing within the square at the intersection of the vertical and horizontal lines thus determined. Then *twelve* squares, equivalent to the twelve different conditions listed in subparagraph *f*, can readily be constructed. They are all shown in Appendix 1, pp. 96–107.

i. When these square tables are examined carefully, certain interesting points are noted. In the first place, the tables may be paired so that one of a pair may serve for enciphering and the other of the pair may serve for deciphering, or vice versa. For example, tables I and II bear this reciprocal relationship to each other; III and IV, V and VI, VII and VIII, IX and X, XI and XII. In the second place, the internal dispositions of the letters, although the tables are derived from the same pair of components, are quite diverse. For example, in table I–B the horizontal sequences are identical with those of Table I–A, but are merely displaced to the right and to the left different intervals according to the successive key letters. Hence this square shows what may be termed a horizontally-displaced, direct symmetry of the cipher component. Vertically, it shows no symmetry, or if there is symmetry, it is not visible.[2] But when Table I–B is more carefully examined, an invisible, or indirect, vertical symmetry may be discerned where at first glance it is not apparent. If one takes any two *columns* of the table, it is found that the interval between the members of any pair of letters in one column is the same as the interval between the members of the homologous pair of letters in the other column, *if the distance is measured on the cipher component*. For example, consider the 2d and 15th columns (headed by L and I, respectively); take the letters P and G in the 2d column, and J and W in the 15th column. The distance between P and G on the cipher component is 7 intervals; the distance between J and W on the same component is *also* 7 intervals. This phenomenon implies a kind of hidden, or latent, or indirect symmetry within the cipher square. In fact, it may be stated that every table which sets forth in systematic fashion the various secondary alphabets derivable by sliding two primary sequences through all points of coincidence to find cipher equivalents must show some kind of symmetry, both horizontally and vertically. The symmetry may be termed *visible* or *direct*, if the sequences of letters in the rows (or columns) are the same throughout and are identical with that of one of the primary components; it may be termed *hidden* or *indirect* if the sequences of letters in the rows or columns are different, apparently not related to either of the components, but are in reality decimations of one of the primary components.

---

[2] It is true that the first column within the table shows the plain-component sequence, but this is merely because the method of finding the equivalents in this case is such that this sequence is bound to appear in that column, since the successive key letters are A, B, C, . . . Z, and this sequence happens to be identical with the plain component in this case. The same is true of Tables V and XI; it is also applicable to the first row of Tables IX and X.

j. When the twelve tables of Appendix 1 are examined in the light of the foregoing remarks, the type of symmetry found in each may be summarized in the following manner:

Table	Horizontal				Vertical			
	Visible or direct		Invisible or indirect		Visible or direct		Invisible or indirect	
	Follows plain component	Follows cipher component	Follows plain component	Follows cipher component	Follows plain component	Follows cipher component	Follows plain component	Follows cipher component
I		x						x
II			x				x	
III		x				x		
IV			x		x			
V		x						x
VI			x				x	
VII	x						x	
VIII	x						x	
IX				x				x
X				x				x
XI			x		x			
XII		x				x		

Of these twelve types of cipher squares, corresponding to the twelve different ways of using a pair of sliding primary components to derive secondary alphabets, the ones best known and most often encountered in cryptographic studies are Tables I–B and II, referred to as being of the Vigenère type; Tables V and VI, referred to as being of the Beaufort type; and Tables IX and X, referred to as being of the Delastelle type. It will be noted that the tables of the Delastelle type show no direct or visible symmetry, either horizontally or vertically and because of this are supposed to yield more security than do any of the other types of tables. But it will presently be shown that the supposed increase in security is more illusory than real.

k. The foregoing facts concerning the various types of quadricular tables generated by diverse methods of using sliding primary components or their equivalent rotating cipher disks will be employed to good advantage, when the studies presently to be undertaken will bring the student to the place where he can comprehend them in the analysis of polyalphabetic systems. But in order not to confuse him with a multiplicity of details which have no direct bearing upon basic principles, one and only one standard method of finding equivalents by means of sliding components will be selected from among the twelve available, as set forth in the preceding subparagraphs. Unless otherwise stated, this method will be the one denoted by the first of the formulae listed in subpar. *f*, *viz:*

$$\theta_{k/2}=\theta_{1/1};\ \theta_{p/1}=\theta_{c/2}$$

Calling the plain component "1" and the cipher component "2", this will mean that the keyletter on the cipher component will be set opposite the index, which will be the first letter of the plain component; the plain-text letter to be enciphered will then be sought on the plain component and its equivalent will be the letter opposite it on the cipher component.

SECTION III

THEORY OF SOLUTION OF REPEATING-KEY SYSTEMS

Paragraph
The three steps in the analysis of repeating-key systems........ 8
First step: finding the length of the period........ 9
General remarks on factoring........ 10
Second step: distributing the cipher text into the component monoalphabets........ 11
Third step: solving the monoalphabetic distributions........ 12

8. The three steps in the analysis of repeating-key systems.—*a.* The method of enciphering according to the principle of the repeating key, or repeating alphabets is adequately explained in Section XI of *Elementary Military Cryptography,* and no further reference need be made at this time. The analysis of a cryptogram of this type, regardless of the kind of cipher alphabets employed, or their method of production, resolves itself into three distinct and successive steps.

(1) Determination of the length of the repeating key, which is the same as the determination of the exact number of alphabets involved in the cryptogram;

(2) Allocation or distribution of the letters of the cipher text into the respective cipher alphabets to which they belong. This is the step which reduces the polyalphabetic text to monoalphabetic terms;

(3) Analysis of the individual monoalphabetic distributions to determine plain-text values of the cipher letters in each distribution or alphabet.

b. The foregoing steps will be treated in the order in which mentioned. The first step may be described briefly as that of *determining the period.* The second step may be described briefly as that of *reduction to monoalphabetic terms.* The third step may be designated as *identification of cipher-text values.*

9. First step: finding the length of the period.—*a.* The determination of the period, that is, the length of the key or the number of cipher alphabets involved in a cryptogram enciphered by the repeating-key method is, as a rule, a relatively simple matter. The cryptogram itself usually manifests externally certain phenomena which are the direct result of the use of a repeating key. The principles involved are, however, so fundamental in cryptanalysis that their elucidation warrants a somewhat detailed treatment. This will be done in connection with a short example of encipherment, shown in Fig. 1.

MESSAGE

THE ARTILLERY BATTALION MARCHING IN THE REAR OF THE ADVANCE GUARD KEEPS ITS COMBAT TRAIN WITH IT INSOFAR AS PRACTICABLE.

[Key: BLUE, using direct standard alphabets]

CIPHER ALPHABETS

	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	(1)	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
	(2)	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
	(3)	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	(4)	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

BLUE	BLUE	BLUE	BLUE
THEA	ARDK	THEA	ARDK
		USYE	BCXO
RTIL	EEPS	RTIL	EEPS
		SECP	FPJW
LERY	ITSC	LERY	ITSC
		MPLC	JEMG
BATT	OMBA	BATT	OMBA
		CLNX	PXVE
ALIO	TTRA	ALIO	TTRA
		BWCS	UELE
NMAR	INWI	NMAR	INWI
		OXUV	JYQM
CHIN	THIT	CHIN	THIT
		DSCR	USCX
GINT	INSO	GINT	INSO
		HTHX	JYMS
HERE	FARA	HERE	FARA
		IPLI	GLLE
AROF	SPRA	AROF	SPRA
		BCIJ	TALE
THEA	CTIC	THEA	CTIC
		USYE	DECG
DVAN	ABLE	DVAN	ABLE
		EGUR	BMFI
CEGU		CEGU	
		DPAY	
a	*a*	*b*	*b*

CRYPTOGRAM

USYES ECPMP LCCLN XBWCS OXUVD SCRHT

HXIPL IBCIJ USYEE GURDP AYBCX OFPJW

JEMGP XVEUE LEJYQ MUSCX JYMSG LLETA

LEDEC GBMFI

FIGURE 1.

b. Regardless of what system is used, identical plain-text letters enciphered by the same cipher alphabet [1] must yield identical cipher letters. Referring to Fig. 1, such a condition is brought about every time that identical plain-text letters happen to be enciphered with the same key-letter, or every time identical plain-text letters fall into the same column in the encipherment.[2] Now since the number of columns or positions with respect to the key is very limited (except in the case of very long key words), and since the repetition of letters is an inevitable condition in plain text, it follows that there will be in a message of fair length many cases where identical plain-text letters *must* fall into the same column. They will thus be enciphered by the same cipher alphabet, resulting, therefore, in the production of many identical letters in the cipher text and these will represent identical letters in the plain text. When identical plain-text polygraphs fall into identical columns the result is the formation of identical cipher-text polygraphs, that is, repetitions of groups of 2, 3, 4, . . . letters are exhibited in the cryptogram. Repetitions of this type will hereafter be called *causal repetitions*, because they are produced by a definite, traceable cause, *viz*, the encipherment of identical letters by the same cipher alphabets.

c. It will also happen, however, that *different* plain-text letters falling in *different* columns will, by mere accident, produce identical cipher letters. Note, for example, in Fig. 1 that in Column 1, R_p becomes S_c and that in Column 2, H_p also becomes S_c. The production of an identical cipher text letter in these two cases (that is, a repetition where the plain-text letters are different and enciphered by different alphabets) is merely fortuitous. It is, in every day language, "a mere coincidence", or "an accident." For this reason repetitions of this type will hereafter be called *accidental repetitions*.

d. A consideration of the phenomenon pointed out in *c* makes it obvious that in polyalphabetic ciphers it is important that the cryptanalyst be able to tell whether the repetitions he finds in a specific case are causal or accidental in their origin, that is, whether they represent actual encipherments of identical plain-text letters by identical keying elements, or mere coincidences brought about purely fortuitously.

e. Now accidental repetitions will, of course, happen fairly frequently with individual letters, but less frequently with digraphs, because in this case the same kind of an "accident" must take place twice in succession. Intuitively one feels that the chances that such a purely fortuitous coincidence will happen two times in succession must be much less than that it will happen every once in a while in the case of single letters. Similarly, intuition makes one feel that the chances of such accidents happening in the case of three or more consecutive letters are still less than in the case of digraphs, decreasing very rapidly as the repetition increases in length.

f. The phenomena of cryptographic repetition may, fortunately, be dealt with statistically, thus taking the matter outside the realm of intuition and putting it on a firm mathematical or objective basis. Moreover, often the statistical analysis will tell the cryptanalyst when he has arranged or rearranged his text properly, that is, when he is approaching or has reached monoalphabeticity in his efforts to reduce polyalphabetic text to its simplest terms. However, in order to preserve continuity of thought it is deemed inadvisable to inject these statistical considerations at this place in the text proper; they have been incorporated in Appendix 2 hereof. The student is advised to study the Appendix very carefully after he has finished reading this section of the text.

g. At this point it will merely be indicated that if a cryptanalyst were to have at hand only the cryptogram of Fig. 1, with the repetitions underlined as below, a statistical study of the number and length of the repetitions within the message (Par. 5 of Appendix 2) would tell him that while some of the digraphic repetitions may be accidental, the chances that they all are accidental are small. In the case of the tetragraphic repetition he would realize that the chances of its being accidental are very small indeed.

---

[1] It is to be understood, of course, that cipher alphabets with single equivalents are meant in this case.

[2] The frequency with which this condition may be *expected* to occur can be definitely calculated. A discussion of this point falls beyond the scope of the present text.

A. USYES ECPMP LCCLN XBWCS OXUVD
B. SCRHT HXIPL IBCIJ USYEE GURDP
C. AYBCX OFPJW JEMGP XVEUE LEJYQ
D. MUSCX JYMSG LLETA LEDEC GBMFI

h. A consideration of the facts therefore leads to but one conclusion, *viz*, that the repetitions exhibited by the cryptogram under investigation are *not accidental* but are *causal* in their origin; and the cause is in this case not difficult to find: repetitions in the plain text were actually enciphered by identical alphabets. In order for this to occur, it was necessary that the tetragraph USYE, for example, fall *both* times in *exactly* the same relative position with respect to the key. Note, for example, that USYE in Fig. 1 represents in both cases the plain-text polygraph THEA. The first time it occurred it fell in positions 1–2–3–4 with respect to the key; the second time it occurred it happened to fall in the very same relative positions, although it might just as well have happened to fall in any of the other three possible relative positions with respect to the key, *viz*, 2–3–4–1, 3–4–1–2, or 4–1–2–3.

i. Lest the student be misled, however, a few more words are necessary on this subject. In the preceding subparagraph the word "happened" was used; this word correctly expresses the idea in mind, because the insertion or deletion of a single plain-text letter between the two occurrences would have thrown the second occurrence one letter forward or backward, respectively, and thus caused the polygraph to be enciphered by a sequence of alphabets such as can no longer produce the cipher polygraph USYE from the plain-text polygraph THEA. On the other hand, the insertion or deletion of this one letter might bring the letters of some other polygraph into similar columns so that some other repetition would be exhibited in case the USYE repetition had thus been suppressed.

j. The encipherment of similar letters by similar cipher alphabets is therefore the *cause* of the production of repetitions in the cipher text in the case of repeating-key ciphers. What principles can be derived from this fact, and how can they be employed in the solution of cryptograms of this type?

k. If a count is made of the number of letters from and including the first USYE to, but not including, the second occurrence of USYE, a total of 40 letters is found to intervene between the two occurrences. This number, 40, must, of course, be an exact multiple of the length of the key. Having the plain-text before one, it is easily seen that it is the 10th multiple; that is, the 4-letter key has repeated itself 10 times between the first and the second occurrence of USYE. It follows, therefore, that if the length of the key were not known, the number 40 could safely be taken to be an exact multiple of the length of the key; in other words, one of the *factors* of the number 40 would be equal to the length of the key. The word "safely" is used in the preceding sentence to mean that the interval 40 applies to a repetition of 4 letters and it has been shown that the chances that this repetition is accidental are small. The factors of 40 are 2, 4, 5, 8, 10, and 20. So far as this single repetition of USYE is concerned, if the length of the key were not known, all that could be said about the latter would be that it is equal to one of these factors. The repetition by itself gives no further indications. How can the exact factor be selected from among a list of several possible factors?

l. Let the intervals between all the repetitions in the cryptogram be listed. They are as follows:

Repetition	Interval	Factors
1st USYE to 2d USYE	40	2, 4, 5, 8, 10, 20.
1st BC to 2d BC	16	2, 4, 8.
1st CX to 2d CX	25	5.
1st EC to 2d EC	88	2, 4, 11, 22, 44.
1st LE to 2d LE	16	2, 4, 8.
2d LE to 3d LE	4	2, 4.
1st LE to 3d LE	20	2, 4, 5, 10.
1st JY to 2d JY	8	2, 4.
1st PL to 2d PL	24	2, 3, 4, 6, 8, 10, 12.
1st SC to 2d SC	52	2, 4, 13, 26.
(1st SY to 2d SY, already included in USYE.)		
(1st US to 2d US, already included in USYE.)		
2d US to 3d US	36	2, 3, 4, 6, 9, 18.
(1st US to 3d US, already included in USYE.)		
(1st YE to 2d YE, already included in USYE.)		

m. Are all these repetitions *causal* repetitions? It can be shown (Appendix 2, par. 4*c*) that the odds against a theory that the USYE repetition is accidental are about 99 to 1 (since the probability for its occurrence is .01). It can also be shown that the odds against a theory that the 10 digraphs which occur two or more times are accidental repetitions are over 4 to 1 (Appendix 2, par. 5*c*); the odds against a theory that the two digraphs which occur 3 times are accidental repetitions are quite large. (Probability is calculated to be about .06.) The chances are very great, therefore, that all or nearly all these repetitions are causal. Certainly the chances against the two occurrences of the tetragraph USYE and the three occurrences of the two different digraphs (LE and US) being accidental are quite high, and it is therefore not astonishing that the intervals between all the various repetitions, except in one case, contain the factors 2 and 4.

n. This means that if the cipher is written out in either 2 columns or 4 columns, all these repetitions (except the CX repetition) would fall into the same columns. From this it follows that the length of the key is either 2 or 4, the latter, on practical grounds, being more probable than the former. Doubts concerning the matter of choosing between a 2-letter and a 4-letter key will be dissolved when the cipher text is distributed into its component uniliteral frequency distributions.

o. The repeated digraph CX in the foregoing message is an accidental repetition, as will be apparent by referring to Fig. 1. Had the message been longer there would have been more such accidental repetitions, but, on the other hand, there would be a proportionately greater number of causal repetitions. This is because the phenomenon of repetition in plain text is so all-pervading.

p. Sometimes it happens that the cryptanalyst quickly notes a repetition of a polygraph of four or more letters, the interval between the first and second occurrences of which has only two factors, of which one is a relatively small number, the other a relatively high incommensurable number. He may therefore assume at once that the length of the key is equal to the smaller factor without searching for additional recurrences upon which to corroborate his assumption. Suppose, for example, that in a relatively short cryptogram the interval between the first and second occurrences of a polygraph of five letters happens to be a number such as 203, the factors of which are 7 and 29. Evidently the number of alphabets may at once be assumed to be 7, unless one is dealing with messages exchanged among correspondents known to use long keys. In the latter case one could assume the number of alphabets to be 29.

q. The foregoing method of determining the period in a polyalphabetic cipher is commonly referred to in the literature as "factoring the intervals between repetitions"; or more often it is simply called "factoring." Because the latter is an apt term and is brief, it will be employed hereafter in this text to designate the process.

10. General remarks on factoring.—*a.* The statement made in Par. 2 with respect to the cyclic phenomena said to be exhibited in cryptograms of the periodic type now becomes clear. The use of a short repeating key produces a periodicity of recurrences or repetitions collectively termed "cyclic phenomena", an analysis of which leads to a determination of the length of the period or cycle, and this gives the length of the key. Only in the case of relatively short cryptograms enciphered by a relatively long key does factoring fail to lead to the correct determination of the number of cipher alphabets in a repeating-key cipher; and of course, the fact that a cryptogram contains repetitions whose factors show constancy is in itself an indication and test of its periodic nature. It also follows that if the cryptogram is not a repeating-key cipher, then factoring will show no definite results, and conversely the fact that it does not yield definite results at once indicates that the cryptogram is not a periodic, repeating-key cipher.

b. There are two cases in which factoring leads to no definite results. One is in the case of monoalphabetic substitution ciphers. Here recurrences are very plentiful as a rule, and the intervals separating these recurrences may be factored, *but the factors will show no constancy;* there will be several factors common to many or most of the recurrences. This in itself is an indication of a monoalphabetic substitution cipher, if the very fact of the presence of many recurrences fails to impress itself upon the inexperienced cryptanalyst. The other case in which the process of factoring is nonsignificant involves certain types of nonperiodic, polyalphabetic ciphers. In certain of these ciphers recurrences of digraphs, trigraphs, and even polygraphs may be plentiful in a long message, but the intervals between such recurrences bear no definite multiple relation to the length of the key, such as in the case of the true periodic, repeating-key cipher, in which the alphabets change with successive letters and repeat themselves over and over again.

c. Factoring is not the only method of determining the length of the period of a periodic, polyalphabetic substitution cipher, although it is by far the most common and easily applied. At this point it will merely be stated that when the message under study is relatively short in comparison with the length of the key, so that there are only a few cycles of cipher text and no long repetitions affording a basis for factoring, there are several other methods available. However, it being deemed inadvisable to interject the data concerning those other methods at this point, they will be explained subsequently. It is desirable at this juncture merely to indicate that methods other than factoring do exist and are used in practical work.

d. Fundamentally, the factoring process is merely a more or less simple mathematical method of studying the phenomena of periodicity in cryptograms. It will usually enable the cryptanalyst to ascertain definitely whether or not a given cryptogram is periodic in nature, and if so, the length of the period, *stated in terms of the cryptographic unit involved.* By the latter statement is meant that the factoring process may be applied not only in analyzing the periodicity manifested by cryptograms in which the plain-text units subjected to cryptographic treatment are monographic in nature (i. e. are single letters) but also in studying the periodicity exhibited by those occasional cryptograms wherein the plain-text units are digraphic, trigraphic, or *n*-graphic in character. The student should bear this point in mind when he comes to the study of substitution systems of the latter sort. However, the present text will deal solely with cases of the former type, wherein the plain-text units subjected to cryptographic treatment are single letters.

11. Second step: distributing the cipher text into the component monoalphabets.—*a*. After the number of cipher alphabets involved in the cryptogram has been ascertained, the next step is to rewrite the message in groups corresponding to the length of the key, or in columnar fashion, whichever is more convenient, and this automatically divides up the text so that the letters belonging to the same cipher alphabet occupy similar positions in the groups, or, if the columnar method is used, fall in the same column. The letters are thus allocated or distributed into the respective cipher alphabets to which they belong. This reduces the polyalphabetic text to monoalphabetic terms.

b. Then separate uniliteral frequency distributions for the thus isolated individual alphabets are compiled. For example, in the case of the cipher on page 13, having determined that four alphabets are involved, and having rewritten the message in four columns, a frequency distribution is made of the letters in Column 1, another is made of the letters in Column 2, and so on for the rest of the columns. *Each of the resulting distributions is therefore a monoalphabetic frequency distribution.* If these distributions do not give the characteristic irregular crest and trough appearance of monoalphabetic frequency distributions, then the analysis which led to the hypothesis as regards the number of alphabets involved is fallacious. In fact, the appearance of these individual distributions may be considered to be an index of the correctness of the factoring process; for theoretically, and practically, the individual distributions constructed upon the *correct* hypothesis will tend to conform more closely to the irregular crest and trough appearacne of a monoalphabetic frequency distribution than will the graphic tables constructed upon an incorrect hypothesis. These individual distributions may also be tested for monoalphabeticity by statistical methods.

12. Third step: solving the monoalphabetic distributions.—The difficulty experienced in analyzing the individual or isolated frequency distributions depends mostly upon the type of cipher alphabets that is used. It is apparent that mixed alphabets may be used just as easily as standard alphabets, and, of course, the cipher letters themselves give no indication as to which is the case. However, just as it was found that in the case of monoalphabetic substitution ciphers, a uniliteral frequency distribution gives clear indications as to whether the cipher alphabet is a standard or a mixed alphabet, by the relative positions and extensions of the crests and troughs in the table, so it is found that in the case of repeating-key ciphers, uniliteral frequency distributions for the isolated or individual alphabets will also give clear indications as to whether these alphabets are standard alphabets or mixed alphabets. Only one or two such frequency distributions are necessary for this determination; if they appear to be standard alphabets, similar distributions can be made for the rest of the alphabets; but if they appear to be mixed alphabets, then it is best to compile triliteral frequency distributions for all the alphabets. The analysis of the values of the cipher letters in each table proceeds along the same lines as in the case of monoalphabetic ciphers. The analysis is more difficult only because of the reduced size of the tables, but if the message be very long, then each frequency distribution will contain a sufficient number of elements to enable a speedy solution to be achieved.

SECTION IV

REPEATING-KEY SYSTEMS WITH STANDARD CIPHER ALPHABETS

Paragraph
Solution by applying principles of frequency 13
Solution by completing the plain-component sequence 14
Solution by the "probable-word method" 15

13. Solution by applying principles of frequency.—*a.* In the light of the foregoing principles, let the following cryptogram be studied:

MESSAGE

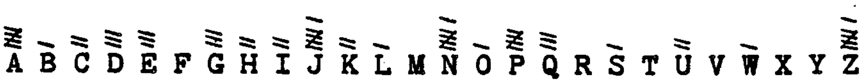
```
        1           2           3           4           5
A.  A U K H Y   J A M K I   Z Y M W M   J M I G X   N F M L X
B.  E T I M I   Z H B H R   A Y M Z M   I L V M E   J K U T G
C.  D P V X K   Q U K H Q   L H V R M   J A Z N G   G Z V X E
D.  N L U F M   P Z J N V   C H U A S   H K Q G K   I P L W P
E.  A J Z X I   G U M T V   D P T E J   E C M Y S   Q Y B A V
F.  A L A H Y   P O E X W   P V N Y E   E Y X E E   U D P X R
G.  B V Z V I   Z I I V O   S P T E G   K U B B R   Q L L X P
H.  W F Q G K   N L L L E   P T I K W   D J Z X I   G O I O I
J.  Z L A M V   K F M W F   N P L Z I   O V V F M   Z K T X G
K.  N L M D F   A A E X I   J L U F M   P Z J N V   C A I G I
L.  U A W P R   N V I W E   J K Z A S   Z L A F M   H S
```

A search for repetitions discloses the following short list with the intervals and factors above 10 omitted (for previous experience may lead to the conclusion that it is unlikely that the cryptogram involves more than 10 alphabets, showing the number of recurrences which it does):

Repetition	Location	Interval	Factors
LUFMPZJNVC	D1, K3	160	2, 4, 5, 8, 10.
JZXIG	E1, H4	90	2, 3, 5, 6, 9, 10.
EJK	B4, L2	215	5.
PTE	E3, G3	50	2, 5, 10.
QGK	D4, H1	85	5.
UKH	A1, C2	55	5.
ZLA	J1, L4	65	5.
AS	D3, L3	175	3, 5, 7,
EJ	B4, L2	115	5.
FM	A5, D1	57	3.
FM	A5, J2	185	5.
FM	J2, J4	12	2, 3, 4, 6.
FM	J4, K3	20	2, 4, 5, 10.
FM	K3, L4	30	2, 3, 5, 6, 10.
JA	A2, C4	60	2, 3, 4, 5, 6, 10.
LA	F1, J1	75	3, 5.
LA	J1, L4	65	5.
LL	G5, H2	10	2, 5.
NL	D1, H2	105	3, 5, 7.
NL	H2, K1	45	3, 5, 9.
VX	C1, C5	20	2, 4, 5, 10.
YM	A3, B3	25	5.

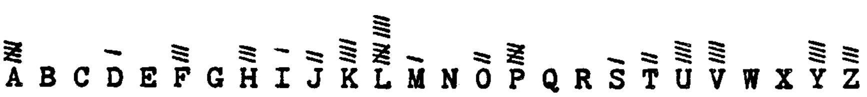
b. The factor 5 appears in all but two cases, each of which involves only a digraph. It seems almost certain that the number of alphabets is five. Since the text already appears in groups of five letters, it is unnecessary to rewrite the message. The next step is to make a uniliteral frequency distribution for Alphabet 1 to see if it can be determined whether or not standard alphabets are involved. It is as follows:

ALPHABET 1



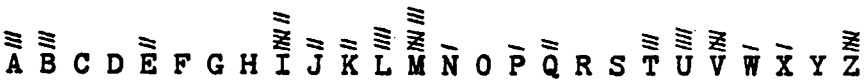
c. Although the indications are not very clear cut, yet if one takes into consideration the small amount of data the assumption of a direct standard alphabet with $W_c=A_p$, is worth further test. Accordingly a similar distribution is made for Alphabet 2.

ALPHABET 2

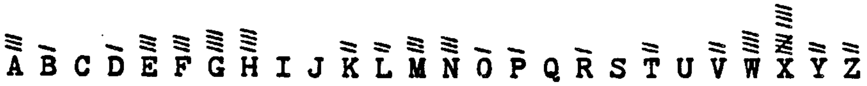


d. There is every indication of a direct standard alphabet, with $H_c=A_p$. Let similar distributions be made for the last three alphabets. They are as follows:

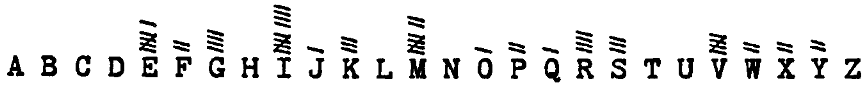
ALPHABET 3



ALPHABET 4



ALPHABET 5



e. After but little experiment it is found that the distributions can best be made to fit the normal when the following values are assumed:

Alphabet 1.......... $A_p=W_c$
Alphabet 2.......... $A_p=H_c$
Alphabet 3.......... $A_p=I_c$
Alphabet 4.......... $A_p=T_c$
Alphabet 5.......... $A_p=E_c$

f. Note the key word given by the successive equivalents of A_p: WHITE. The real proof of the correctness of the analysis is, of course, to test the values of the solved alphabets on the cryptogram. The five complete cipher alphabets are as follows:

	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	2	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
	3	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
	4	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	5	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

FIGURE 2.

g. Applying these values to the first few groups of our message, the following is found:

	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
Cipher........	A U K H Y	J A M K I	Z Y M W M	J M I G X	N F M L X	. . .
Plain...........	E N C O U	N T E R E	D R E D I	N F A N T	R Y E S T	. . .

h. Intelligible text at once results, and the solution can now be completed very quickly. The complete message is as follows:

ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MACHINE GUN COMPANY IN TRUCKS NEAR EMMITSBURG. AM HOLDING MIDDLE CREEK NEAR HILL 543 SOUTHWEST OF FAIRPLAY. WHEN FORCED BACK WILL CONTINUE DELAYING REDS AT MARSH CREEK. HAVE DESTROYED BRIDGES ON MIDDLE CREEK BETWEEN EMMITSBURG–TANEYTOWN ROAD AND RHODES MILL.

i. In the foregoing example (which is typical of the system erroneously attributed, in cryptographic literature, to the French cryptographer Vigenère, although to do him justice, he made no claim of having "invented" it), direct standard alphabets were used, but it is obvious that reversed standard alphabets may be used and the solution accomplished in the same manner. In fact, the now obsolete cipher disk used by the United States Army for a number of years yields exactly this type of cipher, which is also known in the literature as the Beaufort Cipher, and by other names. In fitting the isolated frequency distributions to the normal, the direction of "reading" the crests and troughs is merely reversed.

14. Solution by completing the plain-component sequence.—*a.* There is another method of solving this type of cipher, which is worthwhile explaining, because the underlying principles will be found useful in many cases. It is a modification of the method of solution by completing the plain-component sequence, already explained in *Military Cryptanalysis*, Part I.

b. After all, the individual alphabets of a cipher such as the one just solved are merely direct standard alphabets. It has been seen that monoalphabetic ciphers in which standard cipher alphabets are employed may be solved almost mechanically by completing the plain-component sequence. The plain text reappears on only one generatrix and this generatrix is the same for the whole message. It is easy to pick this generatrix out of all the other generatrices because it is the only one which yields intelligible text. Is it not apparent that if the same process is applied to the cipher letters of the *individual alphabets* of the cipher just solved that the plain-text equivalents of these letters must all reappear on one and the same generatrix? But how will the generatrix which actually contains the plain-text letters be distinguishable from the other generatrices, since these plain-text letters are not consecutive letters in the plain text but only letters separated from one another by a constant interval? The answer is simple. The plain-text generatrix should be distinguishable from the others *because it will show more and a better assortment of high-frequency letters, and can thus be selected by the eye from the whole set of generatrices.* If this is done with all the alphabets in the cryptogram, it will merely be necessary to assemble the letters of the thus selected generatrices in proper order, and the result sould be consecutive letters forming intelligible text.

c. An example will serve to make the process clear. Let the same message be used as before. Factoring showed that it involves five alphabets. Let the first ten cipher letters *in each alphabet* be set down in a horizontal line and let the normal alphabet sequences be completed. Thus:

	ALPHABET 1	ALPHABET 2	ALPHABET 3	ALPHABET 4	ALPHABET 5
1	AJZJNEZAIJ	UAYMFTHYLK	KMMIMIBMVU	HKWGLMHZMT	YIMXXIRMEG
2	BKAKOFABJK	VBZNGUIZML	LNNJNJCNWV	ILXHMNIANU	ZJNYYJSNFH
3	CLBLPGBCKL	WCAOHVJANM	MOOKOKDOXW	JMYINOJBOV	AKOZZKTOGI
4	DMCMQHCDLM	XDBPIWKBON	NPPLPLEPYX	KNZJOPKCPW	BLPAALUPHJ
5	ENDNRIDEMN	YECQJXLCPO	OQQMQMFQZY	LOAKPQLDQX	CMQBBMVQIK
6	FOEOSJEFNO	ZFDRKYMDQP	PRRNRNGRAZ	MPBLQRMERY	DNRCCNWRJL
7	GPFPTKFGOP	AGESLZNERQ	QSSOSOHSBA	NQCMRSNFSZ	EOSDDOXSKM
8	HQGQULGHPQ	BHFTMAOFSR	RTTPTPITCB	ORDNSTOGTA	FPTEEPYTLN
9	IRHRVMHIQR	CIGUNBPGTS	SUUQUQJUDC	PSEOTUPHUB	GQUFFQZUMO
10	JSISWNIJRS	DJHVOCQHUT	TVVRVRKVED	QTFPUVQIVC	HRVGGRAVNP
11	KTJTXOJKST	EKIWPDRIVU	UWWSWSLWFE	RUGQVWRJWD	ISWHHSBWOQ
12	LUKUYPKLTU	FLJXQESJWV	VXXTXTMXGF	SVHRWXSKXE	JTXIITCXPR
13	MVLVZQLMUV	GMKYRFTKXW	WYYUYUNYHG	TWISXYTLYF	KUYJJUDYQS
14	NWMWARMNVW	HNLZSGULYX	XZZVZVOZIH	UXJTYZUMZG	LVZKKVEZRT
15	OXNXBSNOWX	IOMATHVMZY	YAAWAWPAJI	VYKUZAVNAH	MWALLWFASU
16	PYOYCTOPXY	JPNBUIWNAZ	ZBBXBXQBKJ	WZLVABWOBI	NXBMMXGBTV
17	QZPZDUPQYZ	KQOCVJXOBA	ACCYCYRCLK	XAMWBCXPCJ	OYCNNYHCUW
18	RAQAEVQRZA	LRPDWKYPCB	BDDZDZSDML	YBNXCDYQDK	PZDOOZIDVX
19	SBRBFWRSAB	MSQEXLZQDC	CEEAEATENM	ZCOYDEZREL	QAEPPAJEWY
20	TCSCGXSTBC	NTRFYMARED	DFFBFBUFON	ADPZEFASFM	RBFQQBKFXZ
21	UDTDHYTUCD	OUSGZNBSFE	EGGCGCVGPO	BEQAFGBTGN	SCGRRCLGYA
22	VEUEIZUVDE	PVTHAOCTGF	FHHDHDWHQP	CFRCGHCUHO	TDHSSDMHZB
23	WFVFJAVWEF	QWUIBPDUHG	GIIEIEXIRQ	DGSCHIDVIP	UEITTENIAC
24	XGWGKBWXFG	RXVJCQEVIH	HJJFJFYJSR	EHTDIJEWJQ	VFJUUFOJBD
25	YHXHLCXYGH	SYWKDRFWJI	IKKGKGZKTS	FIUEJKFXKR	WGKVVGPKCE
26	ZIYIMDYZHI	TZXLESGXKJ	JLLHLHALUT	GJVFKLGYLS	XHLWWHQLDF

FIGURE 3.

d. If the high-frequency generatrices underlined in Figure 3 are selected and their letters are juxtaposed *in columns* the consecutive letters of intelligible plain text immediately present themselves. Thus:

Selected Generatrices											
	For Alphabet 1, generatrix 5	E	N	D	N	R	I	D	E	M	N
	For Alphabet 2, generatrix 20	N	T	R	F	Y	M	A	R	E	D
	For Alphabet 3, generatrix 19	C	E	E	A	E	A	T	E	N	M
	For Alphabet 4, generatrix 8	O	R	D	N	S	T	O	G	T	A
	For Alphabet 5, generatrix 23	U	E	I	T	T	E	N	I	A	C

Columnar juxtaposition of letters from selected generatrices

1	2	3	4	5
E	N	C	O	U
N	T	E	R	E
D	R	E	D	I
N	F	A	N	T
R	Y	E	S	T
I	M	A	T	E
D	A	T	O	N
E	R	E	G	I
M	E	N	T	A
N	D	M	A	C

FIGURE 4.

Plain text: ENCOUNTERED RED INFANTRY ESTIMATED AT ONE
REGIMENT AND MAC . . . .

e. Solution by this method can thus be achieved without the compilation of any frequency tables whatever and is very quickly attained. The inexperienced cryptanalyst may have difficulty at first in selecting the generatrices which contain the most and the best assortment of high-frequency letters, but with increased practice, a high degree of proficiency is attained. After all it is only a matter of experiment, trial, and error to select and assemble the proper generatrices so as to produce intelligible text.

f. If the letters on the sliding strips were accompanied by numbers representing their relative frequencies in plain text, and these numbers were added *across* each generatrix, then that generatrix with the highest total frequency would *theoretically* always be the plain-text generatrix. Practically it will be among the generatrices which show the first three or four greatest totals. Thus, an entirely mathematical solution for this type of cipher may be applied.

g. If the cipher alphabets are reversed standard alphabets, it is only necessary to convert the cipher letters of each isolated alphabet into their normal, plain-component equivalents and then proceed as in the case of direct standard alphabets..

h. It has been seen how the key word may be discovered in this type of cryptogram. Usually the key is made up of those letters in the successive alphabets whose equivalents are A_p but other conventions are of course possible. Sometimes a key number is used, such as 8–4–7–1–12, which means merely that A_p is represented by the eighth letter from A (in the normal alphabet) in the first cipher alphabet, by the fourth letter from A in the second cipher alphabet, and so on. This modification is known in the literature as the Gronsfeld cipher. However, the method of solution as illustrated above, being independent of the nature of the key, is the same as before.

15. Solution by the "probable-word method."—*a.* The common use of key words in cryptograms such as the foregoing makes possible a method of solution that is simple and can be used where the more detailed method of analysis using frequency distributions or by completing the plain-component sequence is of no avail. In the case of a very short message which may show no recurrences and give no indications as to the number of alphabets involved, this modified method will be found most useful.

b. Briefly, the method consists in assuming the presence of a probable word in the message, and referring to the alphabets to find the key letters applicable when this hypothetical word is assumed to be present in various positions in the cipher text. If the assumed word happens to be correct, and is placed in the correct position in the message, the key letters produced by referring to the alphabets will yield the key word. In the following example it is assumed that reversed standard alphabets are known to be used by the enemy.

MESSAGE

MDSTJ LQCXC KZASA NYYKO LP

c. Extraneous circumstances lead to the assumption of the presence of the word AMMUNITION. One may assume that this word begins the message. Using sliding normal components, one reversed, the other direct, the key letters are ascertained by noting what the successive equivalents of A_p are. Thus:

Cipher	M	D	S	T	J	L	Q	C	X	C
Plain text	A	M	M	U	N	I	T	I	O	N
"Key"	M	P	E	N	W	T	J	K	L	P

The key does not spell any intelligible word. One therefore shifts the assumed word one letter forward and another trial is made.

Cipher	D	S	T	J	L	Q	C	X	C	K
Plain text	A	M	M	U	N	I	T	I	O	N
"Key"	D	E	F	D	Y	Y	V	F	Q	X

This also yields no intelligible key word. One continues to shift the assumed word forward one space at a time until the following point is reached.

Cipher	L	Q	C	X	C	K	Z	A	S	A
Plain text	A	M	M	U	N	I	T	I	O	N
"Key"	L	C	O	R	P	S	S	I	G	N

The key now becomes evident. It is a cyclic permutation of SIGNAL CORPS. It should be clear that since the key word or key phrase repeats itself during the encipherment of such a message, the plain-text word upon whose assumed presence in the message this test is being based may begin to be enciphered at any point in the key, and continue over into its next repetition if it is longer than the key. When this is the case it is merely necessary to shift the latter part of the sequence of key letters to the first part, as in the case noted: LCORPSSIGN is transposed into SIGN . . . LCORPS, and thus SIGNAL CORPS.

d. It will be seen in the foregoing method of solution that the length of the key is of no particular interest or consequence in the steps taken in effecting the solution. The determination of the length and elements of the key comes after the solution rather than before it. In this case the length of the period is seen to be eleven, corresponding to the length of the key (SIGNAL CORPS).

e. The foregoing method is one of the other methods of determining the length of the key (besides factoring), referred to in Par. 10*c*.

f. If the assumption of reversed standard alphabets yields no good results, then direct standard alphabets are assumed and the test made exactly in the same manner. As will be shown subsequently, the method can also be used as a last resort when mixed alphabets are employed.

g. When the assumed word is longer than the key, the sequence of recovered key letters will show a periodicity equal to the length of the key; that is, after a certain number of letters the sequence of key letters will repeat. This phenomenon would be most useful in the case of keys that are not intelligible words but are composed of random letters or figures. Of course, if such a key is longer than the assumed word, this method is of no avail.

h. This method of solution by searching for a word is contingent upon the following circumstances:

(1) That the word whose presence is assumed actually occurs in the message, is properly spelled, and correctly enciphered.

(2) That the sliding components (or equivalent cipher disks or squares) employed in the search for the assumed word are actually the ones which were employed in the encipherment, or are such as to give identical results as the ones which were actually used.

(3) That the pair of enciphering equations used in the test is actually the pair which was employed in the encipherment; or if a cipher square is used in the test, the method of finding equivalents gives results that correspond with those actually obtained in the encipherment. (See par. 9.)

i. The foregoing appears to be quite an array of contingencies and the student may think that on this account the method will often fail. But examining these contingencies one by one, it will be seen that successful application of the method may not be at all rare—after the solution of some messages has disclosed what sort of paraphernalia and methods of employing them are favored by the enemy. From the foregoing remark it is to be inferred that the probable-word method has its greatest usefulness not in an initial solution of a system, but only after successful study of enemy communications by more difficult processes of analysis has told its story to the alert cryptanalyst. Although it is commonly attributed to Bazeries, the French cryptanalyst of 1900, the probable-word method is very old in cryptanalysis and goes back several centuries. Its usefulness in practical work may best be indicated by quoting from a competent observer[1]:

> There is another [method] which is to this first method what the geometric method is to analysis in certain sciences, and, according to the whims of individuals, certain cryptanalysts prefer one to the other. Certain others, incapable of getting the answer with one of the methods in the solution of a difficult problem, conquer it by means of the other, with a disconcerting masterly stroke. This other method is that of the probable word. We may have more or less definite opinions concerning the subject of the cryptogram. We may know something about its date, and the correspondents, who may have been indiscreet in the subject they have treated. On this basis, the hypothesis is made that a certain word probably appears in the text. . . . In certain classes of documents, military or diplomatic telegrams, banking and mining affairs, etc., it is not impossible to make very important assumptions about the presence of certain words in the text. After a cryptanalyst has worked for a long time with the writings of certain correspondents, he gets used to their expressions. He gets a whole load of words to try out; then the changes of key, and sometimes of system, no longer throw into his way the difficulties of an absolutely new study, which might require the analytical method.

To which I am prompted to add the amusing definition of cryptanalysis attributed to a British wag: "All cryptanalysis is divided into two parts: trance-titution and supposition."

---

[1] Givierge, M., *Cours de Cryptographie*, Paris, 1925, p. 30.

SECTION V

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, I

Paragraph
Reason for the use of mixed alphabets........ 16
Interrelated mixed alphabets........ 17
Principles of direct symmetry of position........ 18
Initial steps in the solution of a typical example........ 19
Application of principles of direct symmetry of position........ 20
Subsequent steps in solution........ 21
Completing the solution........ 22
Solution of subsequent messages enciphered by same cipher component........ 23
Summation of relative frequencies as an aid to the selection of the correct generatrices........ 24
Solution by the probable-word method........ 25
Solution when plain component is mixed, the cipher component, the normal........ 26

16. Reason for the use of mixed alphabets.—*a.* It has been seen in the examples considered thus far that the use of several alphabets in the same message does not greatly complicate the analysis of such a cryptogram. There are three reasons why this is so. Firstly, only relatively few alphabets were employed; secondly, these alphabets were employed in a periodic or repeating manner, giving rise to cyclic phenomena in the cryptogram, by means of which the number of alphabets could be determined; and, thirdly, the cipher alphabets were *known* alphabets, by which is meant merely that the sequences of letters in both components of the cipher alphabets were known sequences.

b. In the case of monoalphabetic ciphers it was found that the use of a mixed alphabet delayed the solution to a considerable degree, and it will now be seen that the use of mixed alphabets in polyalphabetic ciphers renders the analysis much more difficult than the use of standard alphabets, but the solution is still fairly easy to achieve.

17. Interrelated mixed alphabets.—*a.* It was stated in Par. 5 that the method of producing the mixed alphabets in a polyalphabetic cipher often affords clues which are of great assistance in the analysis of the cipher alphabets. This is so, of course, only when the cipher alphabets are interrelated secondary alphabets produced by sliding components or their equivalents. Reference is now made to the classification set forth in Par. 6, in connection with the types of alphabets which may be employed in polyalphabetic substitution. It will be seen that thus far only Cases A (1) and (2) have been treated. Case B (1) will now be discussed.

b. Here one of the components, the plain component, is the normal sequence, while the cipher component is a mixed sequence, the various juxtapositions of the two components yielding mixed alphabets. The mixed component may be a systematically-mixed or a random-mixed sequence. If the 25 successive displacements of the mixed component are recorded in separate lines, a symmetrical cipher square such as that shown in Fig. 5 results therefrom. It is identical in form with the square table shown on p. 7, labeled Table I-A.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L
	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E
	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A
	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V
	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N
	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W
	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O
	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R
	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T
	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B
Cipher	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C
	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D
	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F
	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G
	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I
	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J
	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K
	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M
	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P
	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q
	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U
	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X
	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y

FIGURE 5.

c. Such a cipher square may be used in exactly the same manner as the Vigenère square. With the key word BLUE and conforming to the normal enciphering equations $(\Theta_{k/2}=\Theta_{1/1};\ \Theta_{p/1}=\Theta_{c/2})$, the following lines of the square would be used:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L

FIGURE 6a.

These lines would, of course, yield the following cipher alphabets:

(1)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
(2)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
(3)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
(4)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L

FIGURE 6b.

18. Principles of direct symmetry of position.—*a*. It was stated directly above that Fig. 5 is a symmetrical cipher square, by which is meant that the letters in its successive horizontal lines show a *symmetry of position* with respect to one another. They constitute, in reality, one and only one sequence or series of letters, the sequences being merely displaced successively 1, 2, 3, . . . intervals. The symmetry exhibited is obvious and is said to be visible, or *direct*. This fact can be used to good advantage, as has already been alluded to in par. 7*j*.

b. Consider, for example, the pair of letters G_c and V_c in cipher alphabet (1) of Fig. 6*b*. The letter V_c is the 15th letter to the right of G_c. In cipher alphabet (2), V_c is also the 15th letter to the right of G_c, as is the case in each of the four cipher alphabets in Fig. 6*b*, since the *relative* positions they occupy are the same in each horizontal line in Fig. 6*a*, that is, in each of the successive recordings of the cipher component as the latter is slid to the right against the plain or normal component. If, therefore, the relative positions occupied by two letters, θ_1 and θ_2, in such a cipher alphabet, C_1, are known, and if the position of θ_1 in another cipher alphabet, C_2, belonging to the same series is known, then θ_2 may at once be placed into its correct position in C_2. Suppose, for example, that as the result of an analysis based upon considerations of frequency, the following values in four cipher alphabets have been tentatively determined:

(1)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					G										Y					V						
(2)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					N										G					P						
(3)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					L										B					I						
(4)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					W										I					Q						

FIGURE 7a.

c. The cipher components of these four secondary alphabets may, for convenience, be assembled into a cellular structure, hereinafter called a *sequence reconstruction skeleton*, as shown in Fig. 7*b*. Regarding the top line of the reconstruction skeleton in Fig. 7*b* as being common to all four secondary cipher alphabets listed in Fig. 7*a*, the successive lines of the reconstruction skeleton may now be termed cipher alphabets, and may be referred to by the numbers at the left.

Plain		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1					G										Y					V						
	2					N										G					P						
	3					L										B					I						
	4					W										I					Q						

FIGURE 7b.

d. The letter G is common to Alphabets 1 and 2. In Alphabet 2 it is noted that N occupies the 10th position to the left of G, and the letter P occupies the 5th position to the right of G. One may therefore place these letters, N and P, in their proper positions in Alphabet 1, the letter N being placed 10 letters before G, and the letter P, 5 letters after G. Thus:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1					G					P					Y					V	N					

Thus, the values of two new letters in Alphabet 1, viz, $P_c=J_p$, and $N_c=U_p$ have been automatically determined; these values were obtained without any analysis based upon the *frequency* of P_c and N_c. Likewise, in Alphabet 2, the letters Y and V may be inserted in these positions:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2				V	N										G					P					Y	

This gives the new values $V_c=D_p$ and $Y_c=Y_p$ in Alphabet 2. Alphabets 3 and 4 have a common letter I, which permits of the placement of Q and W in Alphabet 3, and of B and L in Alphabet 4.

e. The new values thus found are of course immediately inserted throughout the cryptogram, thus leading to the assumption of further values in the cipher text. This process, *viz, the reconstruction of the primary components*, by the application of the principles of direct symmetry of position to the cells of the reconstruction skeleton, thus facilitates and hastens solution.

f. It must be clearly understood that before the principles of direct symmetry of position can be applied in cases such as the foregoing, *it is necessary that the plain component be a known sequence.* Whether it is the normal sequence or not is immaterial, so long as the sequence is known. Obviously, if the sequence is unknown, symmetry, even if present, cannot be detected by the cryptanalyst because he has no *base* upon which to try out his assumptions for symmetry. In other words, direct symmetry of position is manifested in the illustrative example because the plain component is a known sequence, and not because it is the normal alphabet. The significance of this point will become apparent later on in connection with the problem discussed in Par. 26*b*.

19. Initial steps in the solution of a typical example.—*a.* In the light of the foregoing principles let a typical message now be studied.

MESSAGE

	1	2	3	4	5
A.	Q W B R I	V W Y C A	I S P J L	R B Z E Y	Q W Y E U
B.	L W M G W	I C J C I	M T Z E I	M I B K N	Q W B R I
C.	V W Y I G	B W N B Q	Q C G Q H	I W J K A	G E G X N
D.	I D M R U	V E Z Y G	Q I G V N	C T G Y O	B P D B L
E.	V C G X G	B K Z Z G	I V X C U	N T Z A O	B W F E Q
F.	Q L F C O	M T Y Z T	C C B Y Q	O P D K A	G D G I G
G.	V P W M R	Q I I E W	I C G X G	B L G Q Q	V B G R S
H.	M Y J J Y	Q V F W Y	R W N F L	G X N F W	M C J K X
J.	I D D R U	O P J Q Q	Z R H C N	V W D Y Q	R D G D G
K.	B X D B N	P X F P U	Y X N F G	M P J E L	S A N C D
L.	S E Z Z G	I B E Y U	K D H C A	M B J J F	K I L C J
M.	M F D Z T	C T J R D	M I Y Z Q	A C J R R	S B G Z N
N.	Q Y A H Q	V E D C Q	L X N C L	L V V C S	Q W B I I
P.	I V J R N	W N B R I	V P J E L	T A G D N	I R G Q P
Q.	A T Y E W	C B Y Z T	E V G Q U	V P Y H L	L R Z N Q
R.	X I N B A	I K W J Q	R D Z Y F	K W F Z L	G W F J Q
S.	Q W J Y Q	I B W R X			

b. The principal repetitions of three or more letters have been underlined in the message and the factors (up to 20 only) of the intervals between them are as follows:

QWBRIVWY	45=3, 5, 9, 15.
CGXGB	60=2, 3, 4, 5, 6, 10, 12, 15, 20.
PJEL	95=5, 19.
ZZGI	145=5.
BRIV	285=3, 5, 15, 19.
BRI	45=3, 5, 9, 15.
KAG	75=3, 5, 15.
QRD	165=3, 5, 15.
QWB	45=3, 5, 9, 15.
QWB	275=5, 11.
WIC	130=2, 5, 10, 13.
XNF	45=3, 5, 9, 15.
YZT	225=3, 5, 15.
ZTC	145=3, 5.

The factor 5 is common to all of these repetitions, and there seems to be every indication that five alphabets are involved. Since the message already appears in groups of five letters, it is unnecessary in this case to rewrite it in groups corresponding to the length of the key. The uniliteral frequency distribution for Alphabet 1 is as follows:

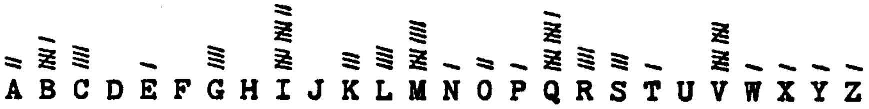


Figure 8.

c. Attempts to fit this distribution to the normal on the basis of a direct or reversed standard alphabet do not give positive results, and it is assumed that mixed alphabets are involved. Individual triliteral frequency distributions are then compiled and are shown in Fig. 9. These tables are similar to those made for single mixed alphabet ciphers, and are made in the same way except that instead of taking the letters one after the other, the letters which belong to the separate alphabets now must be assembled in separate tables. For example, in Alphabet 1, the trigraph QAC means that A occurs in Alphabet 1; Q, its prefix, occurs in Alphabet 5, and C, its suffix, occurs in Alphabet 2. All confusion may be avoided by placing numbers indicating the alphabets in which they belong above the letters, thus: $\overset{5\,1\,2}{\mathrm{QAC}}$.

ALPHABET 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
QC	GW	NF		TV		AE		AS		UD	UW	IT	UT	QP	NX	-W	LB	LA	LA		IW	NN	QI	UX	QR
PT	OP	TG				AD		WC		FI	QX	II		UP		YW	YW	DE			IW				
	GK	TT				LX		HW		-W	LV	OT				NW	QD	RB			UE				
	OW	WB				LW		ND			LR	SY				QC	QD				LC				
	GL							GV				WC				GI					GP				
	GX							WC				GP				QL					QB				
								XD				AB				RI					NW				
								GB				JF				YV					QE				
								IV				DI				NY					IP				
								NR								SW					UP				
								AK								QW									
								QB																	

Alphabet 2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
SN	RZ	IJ	IM	GG	MD			MB		IW	QF		WB		BD		ZH	IP	MZ		IX	QB	GN	MJ	
TG	VG	QG	GG	VZ				QG		BZ	BG				OD		IG		CG		QF	VY	BD	QA	
	IE	VG	ID	SZ				QI							VW		LZ		NZ		LV	QY	PF		
	MJ	CB	RG	VD				KL							OJ				MY		IJ	LM	YN		
	SG	IG	KH					MY							MJ				CJ		EG	QB	LN		
	CY	MJ	RZ					XN							VJ				AY			VY			
	IW	AJ													VY							BN			
																						IJ			
																						BF			
																						RN			
																						VD			
																						QB			
																						KF			
																						GF			
																						QJ			

Alphabet 3

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
YH	WR		PB	BY	WE	CQ	RC	IE	CC		IC	WG	WB		SJ						VC	PM	VC	WC	BE
	IK		PK		LC	EX	DC		WK			DR	WF									KJ		WE	TE
	WR		DR		VW	IV			YJ				XF									BR		WI	EY
	CY		WY		XP	TY			CK				XF											TZ	KZ
	WI		XB		WZ	CX			PQ				AC											IZ	TA
	NR		FZ		WJ	DI			PE				XC											TE	EZ
			EC			CX			BJ				IB											BZ	RN
						LQ			TR															PH	DY
						BR			CR																
						DD			VR																
						BZ			PE																
						AD			WY																
						RQ																			
						VQ																			

Alphabet 4

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ZO	NQ	YA	GG	ZY	NL	MW	AQ	YG	PL	BN		WR	ZQ		FU	GH	BI				GN	FY	GN	ZG	ZG
	DL	JI	GN	YU	NW		YL	GG	JY	JA						GQ	BI						GG	GO	YT
	DN	XU		ZI	NG			BI	JF	DA						JQ	MU						GG	BQ	ZG
	NA	FO		FQ					WQ	JX						GP	GS							DQ	DT
		HN		IW					FQ							GU	DU							EU	YQ
		ND		JL													JD							ZF	GN
		HA		JL													JR							JQ	YT
		LJ		YW													JN								FL
		DQ															BI								
		NL															WX								
		VS																							

ALPHABET 5

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
CI			CS		JK	IB	QI	RV	CM		JR		KQ	YB	QA	BQ	MQ	RM	ZC	EL		GI	KI	EQ	
KG			RM		YK	YQ		CM			BV		XI	AB		EQ	RS	CQ	ZC	RV		EI	R–	JQ	
KG						XB		EM			FG		VC	CM		YO			ZE	CN		FM		WR	
CM						ZI		RV			ES		CV			QV				RO		EC			
BI						IV		II			CL		BP			QZ				PY					
						XB		RV			ET		ZQ			YR				YK					
						DB					HL		RW			ZA				QV					
						FM					ZG		DI			HV									
						ZI										CL									
																NX									
																JR									
																JQ									
																YI									

Condensed table of repetitions

1–2–3–4–5–1–2–3
Q W B R I V W Y–2

2–3–4–5–1
C G X G B–2

2–3–4–1
P J E L–2

3–4–5–1
B–R–I–V
Z–Z–G–I–2

1–2–3
Q W B–3
V W Y–2

2–3–4
C G X–2
P J E–2
W B R–2
X N F–2

3–4–5
B R I–3
G X G–2
J E L–2
Y Z T–2
Z Z G–2

4–5–1
K A G–2
X G B–2
Z G I–2
Z T C–2
R I V–3

5–1–2
I V W–2
Q R D–2
W I C–2

1–2
Q W–5
V P–3
V W–3

2–3
C G–3
C J–3
P J–3
W B–3
W F–3
W Y–3
X N–3

3–4
B R–3
G Q–4
G X–3
J R–3
N F–3
Y Z–3

4–5
R I–3
Y Q–3
Z T–3

5–1
G B–4
I V–3
Q Q–3

FIGURE 9.

d. One now proceeds to analyze each alphabet distribution, in an endeavor to establish identifications of cipher equivalents. First, of course, attempts should be made to separate the vowels from the consonants in each alphabet, using the same test as in the case of a single mixed-alphabet cipher. There seems to be no doubt about the equivalent of E_p in each alphabet:

$$E=\overset{1}{I}_c,\overset{2}{W}_c,\overset{3}{G}_c,\overset{4}{C}_c,\overset{5}{Q}_c$$

e. The letters of greatest frequency in Alphabet 1 are I, M, Q, V, B, G, L, R, S, and C. I_c has already been assumed to be E_p. If $\overset{2}{W}_c$ and $\overset{5}{Q}_c=E_p$, then one should be able to distinguish the vowels from the consonants among the letters M, Q, V, B, G, L, R, S, and C by examining the prefixes of $\overset{2}{W}_c$, and the suffixes of $\overset{5}{Q}_c$. The prefixes and suffixes of these letters, as shown by the triliteral frequency distributions, are these:

Prefixes of $\overset{2}{W}_c$ $(=\overset{2}{E}_p)$	Suffixes of $\overset{5}{Q}_c$ $(=\overset{5}{E}_p)$
$\underset{\not\equiv}{Q}\ \underset{-}{G}\ \underset{-}{K}\ \underset{\equiv}{V}\ \underset{-}{R}\ \underset{=}{B}\ \underset{-}{I}\ \underset{-}{L}$	$\overset{-}{I}\ \overset{\equiv}{Q}\ \overset{=}{R}\ \overset{-}{X}\ \overset{-}{L}\ \overset{=}{V}\ \overset{-}{A}\ \overset{-}{Z}\ \overset{-}{O}$

f. Consider now the letter $\overset{1}{M}_c$; it does not occur either as a prefix of $\overset{2}{W}_c$, or as a suffix of $\overset{5}{Q}_c$. Hence it is most probably a vowel, and on account of its high frequency it may be assumed to be O_p. On the other hand, note that $\overset{1}{Q}_c$ occurs five times as a prefix of $\overset{2}{W}_c$ and three times as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, most probably R_p, for it would give the digraph ER $(=\overset{5}{Q}\overset{1}{Q}_c)$ as occurring three times and RE $(=\overset{1}{Q}\overset{2}{W}_c)$ as occurring five times.

g. The letter $\overset{1}{V}_c$ occurs three times as a prefix of $\overset{2}{W}_c$ and twice as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, and on account of its frequency, let it be assumed to be T_p. The letter $\overset{1}{B}_c$ occurs twice as a prefix of $\overset{2}{W}_c$ but not as a suffix of $\overset{5}{Q}_c$. Its frequency is only medium, and it is probably a consonant. In fact, the twice repeated digraph $\overset{1}{B}\overset{2}{W}_c$ is once a part of the trigraph $\overset{5}{G}\overset{1}{B}\overset{2}{W}$, and $\overset{5}{G}_c$, the letter of second highest frequency in Alphabet 5, looks excellent for T_p. Might not the trigraph $\overset{5}{G}\overset{1}{B}\overset{2}{W}$ be THE? It will be well to keep this possibility in mind.

h. The letter $\overset{1}{G}_c$ occurs only once as a prefix of $\overset{2}{W}_c$ and does not occur as a suffix of $\overset{5}{Q}_c$. It may be a vowel, but one can not be sure. The letter $\overset{1}{L}_c$ occurs once as a prefix of $\overset{2}{W}_c$ and once as a suffix of $\overset{5}{Q}_c$. It may be considered to be a consonant. $\overset{1}{R}_c$ occurs once as a prefix of $\overset{2}{W}_c$, and twice as a suffix of $\overset{5}{Q}_c$, and is certainly a consonant. Neither the letter $\overset{1}{S}_c$ nor the letter $\overset{1}{C}_c$ occurs as a prefix of $\overset{2}{W}_c$ or as a suffix of $\overset{5}{Q}_c$; both would seem to be vowels, but a study of the prefixes and suffixes of these letters lends more weight to the assumption that $\overset{1}{C}_c$ is a vowel than that $\overset{1}{S}_c$ is a vowel. For all the prefixes of C, viz, $\overset{5}{N}$, $\overset{5}{T}$, and $\overset{5}{W}$, are in subsequent analysis of Alphabet 5 classified as consonants, as are likewise its suffixes, viz, T, C, and B in Alphabet 2. On the other hand, only one prefix, $\overset{5}{L}_c$, and one suffix, $\overset{2}{B}_c$, of $\overset{1}{S}_c$ are later classified as consonants. Since vowels are more often associated with consonants than with other vowels, it would seem that $\overset{1}{C}_{e}$ is more likely to be a vowel than $\overset{1}{S}_{e}$. At any rate $\overset{1}{C}_{e}$ is assumed to be a vowel, for the present, leaving $\overset{1}{S}_{e}$ unclassified.

i. Going through the same steps with the remaining alphabets, the following results are obtained:

Alphabet	Consonants	Vowels
1	Q, V, B, L, R, G?	I, M, C.
2	B, C, D, T.	W, P, I.
3	J, N, D, Y, F.	G, Z.
4	Y, Z, J. Q.	C, E?, R?, B?
5	G, N, A, I, W, L, T.	Q, U.

20. Application of principles of direct symmetry of position.—*a*. The next step is to try to determine a few values in each alphabet. In Alphabet 1, from the foregoing analysis, the following data are on hand:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	C?				I				C?						M			Q		V						

Let the values of E_p already assumed in the remaining alphabets, be set down in a reconstruction skeleton, as follows:

Plain		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1	C?				I				C?						M			Q		V						
	2					W																					
	3					G																					
	4					C																					
	5					Q																					

FIGURE 10.

b. It is seen that by good fortune the letter Q is common to Alphabets 1 and 5, and the letter C is common to Alphabets 1 and 4. If it is assumed that one is dealing with a case in which a mixed component is sliding against the normal component, one can apply the principles of direct symmetry of position to these alphabets, as outlined in Par. 18. For example, one may insert the following values in Alphabet 5:

Plain		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1	C?				I				C?						M			Q		V						
	5		M			Q		V							C?				I				C?				

FIGURE 11.

c. The process at once gives three definite values: $\overset{5}{M}_c=B_p$, $\overset{5}{V}_c=G_p$, $\overset{5}{I}_c=R_p$. Let these deduced values be substantiated by referring to the frequency distribution. Since B and G are normally low or medium frequency letters in plain text, one should find that M_c and V_c, their hypothetical equivalents in Alphabet 5, should have low frequencies. As a matter of fact, they do not appear in this alphabet, which thus far corroborates the assumption. On the other hand, since $\overset{5}{I}_c=R_p$, if the values derived from symmetry of position are correct, $\overset{5}{I}_c$ should be of high frequency, and reference to the distribution shows that I_c is of high frequency. The position of C is doubtful; it belongs either under N_p or V_p. If the former is correct, then the frequency of $\overset{5}{C}_c$ should be high, for it would equal N_p; if the latter is correct, then its frequency should be low, for it would equal V_c. As a matter of fact, $\overset{5}{C}_c$ does not occur, and it must be concluded that it belongs under V_p. This in turn settles the value of $\overset{1}{C}_c$, for it must now be placed definitely under I_p and removed from beneath A_p.

d. The definite placement of C now permits the insertion of new values in Alphabet 4, and one now has the following:

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1					I				C						M			Q		V						
	2					W																					
	3					G																					
	4	I				C						M			Q		V										
	5		M			Q		V											I				C				

FIGURE 12.

21. Subsequent steps in solution.—*a.* It is high time that the thus far deduced values, as recorded in the reconstruction skeleton, be inserted in the cipher text, for by this time it must seem that the analysis has certainly gone too far upon unproved hypotheses. The following results are obtained:

MESSAGE

```
       1        2        3        4        5
A.  Q W B R I  V W Y C A  I S P J L  R B Z E Y  Q W Y E U
    R E     R  T E     E  E                     R E

B.  L W M G W  I C J C I  M T Z E I  M I B K N  Q W B R I
      E        E     E R  O       R  O          R E     R

C.  V W Y I G  B W N B Q  Q C G Q H  I W J K A  G E G X N
    T E   A      E     E  R   E N    E E            E
```

```
D.  IDMRU  VEZYG  QIGVN  CTGYO  BPDBL
    E      T      R EP   I E

E.  VCGXG  BKZZG  IVXCU  NTZAO  BWFEQ
    T E           E  E           E  E

F.  QLFCO  MTYZT  CCBYQ  OPDKA  GDGIG
    R  E   O      I   E           EA

G.  VPWMR  QITEW  ICGXG  BLGQQ  VBGRS
    T  K   R      E E      ENE  T E

H.  MYJJY  QVFWY  RWNFL  GXNFW  MCJKX
    O      R       E            O

J.  IDDRU  OPJQQ  ZRHCN  VWDYQ  RDGDG
    E         NE     E   TE  E    E

K.  BXDBN  PXFPU  YXNFG  MPJEL  SANCD
                         O         E

L.  SEZZG  IBEYU  KDHCA  MBJJF  KILCJ
           E         E   O         E

M.  MFDZT  CTJRD  MIYZQ  ACJRR  SBGZN
    O      I      O   E           E

N.  QYAHQ  VEDCQ  LXNCL  LVVCS  QWBII
    R   E  T  EE     E      E   RE AR

P.  IVJRN  WNBRI  VPJEL  TAGDN  IRGQP
    E          R  T        E    E EN

Q.  ATYEW  CBYZT  EVGQU  VPYHL  LRZNQ
           I        EN   T          E

R.  XINBA  IKWJQ  RDZYF  KWFZL  GWFJQ
           E   E          E      E  E

S.  QWJYQ  IBWRX
    RE  E  E
```

b. The combinations given are excellent throughout and no inconsistencies appear. Note the trigraph $\overset{123}{QWB}$, which is repeated in the following polygraphs (underlined in the foregoing text):

1	2	3	4	5	1		5	1	2	3	4	5	1
Q	W	B	R	I	V	. . .	S	Q	W	B	I	I	I
R	E			R	T	. . .		R	E	.	A	R	E

c. The letter $\overset{3}{B}_c$ is common to both polygraphs, and a little imagination will lead to the assumption of the value $\overset{3}{B}_c=P_p$, yielding the following:

1	2	3	4	5	1		5	1	2	3	4	5	1
Q	W	B	R	I	V	. . .	S	Q	W	B	I	I	I
R	E	P	O	R	T	. . .	P	R	E	P	A	R	E

d. Note also (in F5) the polygraph $\begin{array}{cccccc}4&5&1&2&3&4\\ I&G&V&P&W&M\\ A&&T&&&K\end{array}$, which looks like the word ATTACK. The frequency distributions are consulted to see whether the frequencies given for $\overset{5}{G}_c$ and $\overset{2}{P}_c$ are high enough for T_p and A_p, respectively, and also whether the frequency of $\overset{3}{W}_c$ is good enough for C_p; it is noted that they are excellent. Moreover, the digraph $\overset{51}{GB}_c$, which occurs four times, looks like TH, thus making $\overset{1}{B}_c=H_p$. Does the insertion of these four new values in our diagram of alphabets bring forth any inconsistencies? The insertion of the value $\overset{2}{P}_c=A_p$ and $\overset{1}{B}_c=H_p$ gives no indications either way, since neither letter has yet been located in any of the other alphabets. The insertion of the value $\overset{5}{G}_c=T_p$ gives a value common to Alphabets 3 and 5, for the value $\overset{3}{G}_c=E_p$ was assumed long ago. Unfortunately an inconsistency is found here. The letter I has been placed two letters to the left of G in the mixed component, and has given good results in Alphabets 1 and 5; if the value $\overset{3}{W}_c=C_p$ (obtained above from the assumption of the word ATTACK) is correct, then W, and not I, should be the second letter to the left of G. Which shall be retained? There has been so far nothing to establish the value of $\overset{3}{G}_c=E_p$; this value was assumed from frequency considerations solely. Perhaps it is wrong. It certainly behaves like a vowel, and one may see what happens when one changes its value to O_p. The following placements in the reconstruction skeleton result from the analysis, when only two or three new values have been added as a result of the clues afforded by the deductions:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher 1			S		I		G	B	C						M		P	Q	R	V	W					
2	P	Q	R	V	W								S		I		G	B	C						M	
3	R	V	W								S		I		G	B	C						M		P	Q
4	I		G	B	C						M		P	Q	R	V	W								S	
5		M		P	Q	R	V	W								S		I		G	B	C				

FIGURE 12a.

e. Many new values are produced, and these are inserted throughout the message, yielding the following:

```
        1           2           3           4           5
A.  Q W B R I   V W Y C A   I S P J L   R B Z E Y   Q W Y E U
    R E P O R   T E   E     E M Y       S R         R E

B.  L W M G W   I C J C I   M T Z E I   M I B K N   Q W B R I
      E W C H   E S   E R   O       R   O O P       R E P O R

C.  V W Y I G   B W N B Q   Q C G Q H   I W J K A   G E G X N
    T E   A T   H E   D E   R S O N     E E         G   O

D.  I D M R U   V E Z Y G   Q I G V N   C T G Y O   B P D B L
    E   W O     T       T   R O O P     I   O       H A   D

E.  V C G X G   B K Z Z G   I V X C U   N T Z A O   B W F E Q
    T S O   T   H       T   E D   E                 H E     E

F.  Q L F C O   M T Y Z T   C C B Y Q   O P D K A   G D G I G
    R     E     O           I S P   E     A         G   O A T

G.  V P W M R   Q I I E W   I C G X G   B L G Q Q   V B G R S
    T A C K F   R O M   H   E S O   T   H   O N E   T R O O P

H.  M Y J J Y   Q V F W Y   R W N F L   G X N F W   M C J K X
    O           R D   Q     S E         G       H   O S

J.  I D D R U   O P J Q Q   Z R H C N   V W D Y Q   R D G D G
    E     O       A   N E     C   E     T E     E   S   O   T

K.  B X D B N   P X F P U   Y X N F G   M P J E L   S A N C D
    H     D     Q     M             T   O A         C     E

L.  S E Z Z G   I B E Y U   K D H C A   M B J J F   K I L C J
    C       T   E R               E     O R           O   E

M.  M F D Z T   C T J R D   M I Y Z Q   A C J R R   S B G Z N
    O           I     O     O O     E     S   O F   C R O

N.  Q Y A H Q   V E D C Q   L X N C L   L V V C S   Q W B I I
    R       E   T     E E         E       D B E P   R E P A R

P.  I V J R N   W N B R I   V P J E L   T A G D N   I R G Q P
    E D   O     U   P O R   T A             O       E C O N D

Q.  A T Y E W   C B Y Z T   E V G Q U   V P Y H L   L R Z N Q
            H   I R           D O N     T A           C     E

R.  X I N B A   I K W J Q   R D Z Y F   K W F Z L   G W F J Q
      O   D     E       E   S             E         G E     E

S.  Q W J Y Q   I B W R X
    R E     E   E R   O
```

22. Completing the solution.—*a*. Completion of solution is now a very easy matter. The mixed component is finally found to be the following sequence, based upon the word EXHAUSTING:

E X H A U S T I N G B C D F J K L M O P Q R V W Y Z

and the completely reconstructed skeleton of the cipher square is shown in Fig. 13*b*.

Plain		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	1	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H
	2	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O
Cipher	3	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q
	4	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T
	5	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K

FIGURE 13b.

b. Note that the successive equivalents of A_p spell the word APRIL, which is the key for the message. The plain-text message is as follows:

REPORTED ENEMY HAS RETIRED TO NEWCHESTER. ONE TROOP IS REPORTED AT HENDERSON MEETING HOUSE: TWO OTHER TROOPS IN ORCHARD AT SOUTHWEST EDGE OF NEWCHESTER. 2D SQ IS PREPARING TO ATTACK FROM THE SOUTH. ONE TROOP OF 3D SQ IS ENGAGING HOSTILE TROOP AT NEWCHESTER. REST OF 3D SQ IS MOVING TO ATTACK NEWCHESTER FROM THE NORTH. MOVE YOUR SQ INTO WOODS EAST OF CROSSROAD 539 AND BE PREPARED TO SUPPORT ATTACK OF 2D AND 3D SQ. DO NOT ADVANCE BEYOND NEWCHESTER. MESSAGES HERE.

TREER,
COL.

c. The preceding case is a good example of the value of the principles of direct symmetry of position when applied properly to a cryptogram enciphered by the sliding of a mixed component against the normal. The cryptanalyst starts off with only a very limited number of assumptions and builds up many new values as a result of the placement of the few original values in the reconstruction skeleton.

23. Solution of subsequent messages enciphered by the same cipher component.—*a*. *Preliminary remarks*.—Let it be supposed that the correspondents are using the same basic or primary component but with different key words for other messages. Can the knowledge of the sequence of letters in the reconstructed primary component be used to solve the subsequent messages? It has been shown that in the case of a monoalphabetic cipher in which a mixed alphabet was used, the process of completing the plain component could be applied to solve subsequent messages in which the same cipher component was used, even though the cipher component was set at a different key letter. A modification of the procedure used in that case can be used in this case, where a plurality of cipher alphabets based upon a sliding primary component is used.

b. The message.—Let it be supposed that the following message passing between the same two correspondents as in the preceding message has been intercepted:

MESSAGE

SFDZR YRRKX MIWLL AQRLU RQFRT IJQKF XUWBS MDJZK
MICQC UDPTV TYRNH TRORV BQLTI QBNPR RTUHD PTIVE
RMGQN LRATQ PLUKR KGRZF JCMGP IHSMR GQRFX BCABA
OEMTL PCXJM RGQSZ VB

c. Factoring and conversion into plain component equivalents.—The presence of a repetition of a four-letter polygraph whose interval is 21 letters suggests a key word of seven letters. There are very few other repetitions, and this is to be expected in a short message with a key of such length.

1	2	3	4	5	6	7
S	F	D	Z	R	Y	R
R	K	X	M	I	W	L
L	A	Q	R	L	U	R
Q	F	R	T	I	J	Q
K	F	X	U	W	B	S
M	D	J	Z	K	M	I
C	Q	C	U	D	P	T
V	T	Y	R	N	H	T
R	O	R	V	B	Q	L
T	I	Q	B	N	P	R
R	T	U	H	D	P	T
I	V	E	R	M	G	Q
N	L	R	A	T	Q	P
L	U	K	R	K	G	R
Z	F	J	C	M	G	P
I	H	S	M	R	G	Q
R	F	X	B	C	A	B
A	O	E	M	T	L	P
C	X	J	M	R	G	Q
S	Z	V	B			

FIGURE 14.

d. Transcription into periods.—Let the message be written in groups of seven letters, in columnar fashion, as shown in Fig. 14. The letters in each column belong to a single alphabet. Let the letters in each column be converted into their plain-component equivalents by setting the reconstructed cipher component against the normal alphabet at any arbitrarily selected point, for example, that shown below:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z

The columns of equivalents are now as shown in Fig. 15.

1	2	3	4	5	6	7
F	N	M	Z	V	Y	V
V	P	B	R	H	X	Q
Q	D	U	V	Q	E	V
U	N	V	G	H	O	U
P	N	B	E	X	K	F
R	M	O	Z	P	R	H
L	U	L	E	M	T	G
W	G	Y	V	I	C	G
V	S	V	W	K	U	Q
G	H	U	K	I	T	V
V	G	E	C	M	T	G
H	W	A	V	R	J	U
I	Q	V	D	G	U	T
Q	E	P	V	P	J	V
Z	N	O	L	R	J	T
H	C	F	R	V	J	U
V	N	B	K	L	D	K
D	S	A	R	G	Q	T
L	B	O	R	V	J	U
F	Z	W	K			

FIGURE 15.

e. Examination and selection of generatrices.—It has been shown that in the case of a monoalphabetic cipher it was merely necessary to complete the normal alphabet sequence beneath the plain-component equivalents and the plain text all reappeared on one generatrix. It was also found that in the case of a multiple-alphabet cipher involving standard alphabets, the plain-text equivalents of each alphabet reappeared on the same generatrix, and it was necessary only to combine the proper generatrices in order to produce the plain text of the message. In the case at hand both processes are combined: the normal alphabet sequence is continued beneath the letters of each column and then the generatrices are combined to produce the plain text. The completely developed generatrix diagrams for the first two columns are as follows (Fig. 16):

	COLUMN 1		COLUMN 2
	FVQUPRLWVGVHIQZHVDLF		NPDNNMUGSHGWQENCNSBZ
1	GWRVQSMXWHWIJRAIWEMG	1	OQEOONVHTIHXRFODOTCA
2	HXSWRTNYXIXJKSBJXFNH	2	PRFPPOWIUJIYSGPEPUDB
3	IYTXSUOZYJYKLTCKYGOI	3	QSGQQPXJVKJZTHQFQVEC
4	JZUYTVPAZKZLMUDLZHPJ	4	RTHRRQYKWLKAUIRGRWFD
5	KAVZUWQBALAMNVEMAIQK	5	SUISSRZLXMLBVJSHSXGE
6	LBWAVXRCBMBNOWFNBJRL	6	TVJTTSAMYNMCWKTITYHF
7	MCXBWYSDCNCOPXGOCKSM	7	UWKUUTBNZONDXLUJUZIG
8	NDYCXZTEDODPQYHPDLTN	8	VXLVVUCOAPOEYMVKVAJH
9	OEZDYAUFEPEQRZIQEMUO	9	WYMWWVDPBQPFZNWLWBKI
10	PFAEZBVGFQFRSAJRFNVP	10	XZNXXWEQCRQGAOXMXCLJ
11	QGBFACWHGRGSTBKSGOWQ	11	YAOYYXFRDSRHBPYNYDMK
12	RHCGBDXIHSHTUCLTHPXR	12	ZBPZZYGSETSICQZOZENL
13	SIDHCEYJITIUVDMUIQYS	13	ACQAAZHTFUTJDRAPAFOM
14	TJEIDFZKJUJVWENVJRZT	14	BDRBBAIUGVUKESBQBGPN
15	UKFJEGALKVKWXFOWKSAU	15	CESCCBJVHWVLFTCRCHQO
16	VLGKFHBMLWLXYGPXLTBV	16	DFTDDCKWIXWMGUDSDIRP
17	WMHLGICNMXMYZHQYMUCW	17	EGUEEDLXJYXNHVETEJSQ
18	XNIMHJDONYNZAIRZNVDX	18	FHVFFEMYKZYOIWFUFKTR
19	YOJNIKEPOZOABJSAOWEY	19	GIWGGFNZLAZPJXGVGLUS
20	ZPKOJLFQPAPBCKTBPXFZ	20	HJXHHGOAMBAQKYHWHMVT
21	AQLPKMGRQBQCDLUCQYGA	21	IKYIIHPBNCBRLZIXINWU
22	BRMQLNHSRCRDEMVDRZHB	22	JLZJJIQCODCSMAJYJOXV
23	CSNRMOITSDSEFNWESAIC	23	KMAKKJRDPEDTNBKZKPYW
24	DTOSNPJUTETFGOXFTBJD	24	LNBLLKSEQFEUOCLALQZX
25	EUPTOQKVUFUGHPYGUCKE	25	MOCMMLTFRGFVPDMBMRAY

FIGURE 16.

1	2
C	O
S	Q
N	E
R	O
M	O
O	N
I	V
T	H
S	T
D	I
S	H
E	X
F	R
N	F
W	O
E	D
S	O
A	T
I	C
C	A

FIGURE 17a.

f. Combining the selected generatrices.—After some experimenting with these generatrices the 23d generatrix of Column 1 and the 1st of Column 2, which yield the digraphs shown in Fig. 17*a*, are combined. The generatrices of the subsequent columns are examined to select those which may be added to these already selected in order to build up the plain text. The results are shown in Fig. 17*b*. This process is a very valuable aid in the solution of messages after the primary component has been recovered as a result of the longer and more detailed analysis of the frequency distributions of the first message intercepted. Very often a short message can be solved in no other way than the one shown, if the primary component is completely known.

g. Recovery of the key.—It may be of interest to find the key word for the message. Assuming that enciphering method number 1 (see Par. 7*f*, page 6) were known to be employed, all that is necessary is to set the mixed component of the cipher alphabet underneath the plain component so as to produce the cipher letter indicated as the equivalent of any given plain-text letter in each of the alphabets. For example, in the first alphabet it is noted that $C_p=S_c$. Adjust the two components under each other so as to bring S of the cipher component beneath C of the plain component, thus:

1	2	3	4	5	6	7
C	O	F	I	R	S	T
S	Q	U	A	D	R	O
N	E	N	E	M	Y	T
R	O	O	P	D	I	S
M	O	U	N	T	E	D
O	N	H	I	L	L	F
I	V	E	N	I	N	E
T	H	R	E	E	W	E
S	T	O	F	G	O	O
D	I	N	T	E	N	T
S	H	X	L	I	N	E
E	X	T	E	N	D	S
F	R	O	M	C	O	R
N	F	I	E	L	D	T
W	O	H	U	N	D	R
E	D	Y	A	R	D	S
S	O	U	T	H	X	I
A	T	T	A	C	K	R
I	C	H	A	R	D	S
C	A	P	T			

FIGURE 17b.

```
Plain........ ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher.......                        EXHAUSTINGBCDFJKLMOPQRVWYZ
```

It is noted that $A_p=A_c$. Hence, the first letter of the key word to the message is A. The 2d, 3d, 4th, . . . 7th key letters are found in exactly the same manner, and the following is obtained:

When C O F I R S T equals
S F D Z R Y R then A_p successively equals
A Z I M U T H

24. Summation of relative frequencies as an aid to the selection of the correct generatrices.—*a*. In the foregoing example, under subparagraph *f*, there occurs this phrase: "After some experimenting with these generatrices . . ." By this was meant, of course, that the selection of the correct initial pair of generatrices of plain-text equivalents is in this process a matter of trial and error. The test of "correctness" is whether, when juxtaposed, the two generatrices so selected yield "good" digraphs, that is, high-frequency digraphs such as occur in normal plain text. In his early efforts the student may have some difficulty in selecting, merely by ocular examination, the most likely generatrices to try. There may be in each diagram several generatrices which contain good assortments of high-frequency letters, and the number of trials of combinations of generatrices may be quite large. Perhaps a simple mathematical method may be of assistance in the process.

b. Suppose, in Fig. 16, that each letter were accompanied by a number which corresponds to its relative frequency in normal English telegraphic text. Then, by adding the numbers along each *horizontal* line, the totals thus obtained will serve as relative numerical measures of the frequency values of the respective generatrices. Theoretically, the generatrix with the greatest value will be the correct generatrix because its total will represent the sum of the individual values of the actual plaintext letters. In actual practice, of course, the generatrix with the greatest value may not be the correct one, but the correct one will certainly be among the three or four generatrices with the largest values. Thus, the number of trials may be greatly reduced, in the attempt to put together the correct generatrices.

c. Using the preceding message as an example, note the respective generatrix values in Fig. 18. The frequency values of the respective letters shown in the figure are based upon the normal distribution for War Department telegraphic text (see Table 3, Appendix 1, Military Cryptanalysis, Part I).

Column 1

Generatrix																					Frequency value
0	F	V	Q	U	P	R	L	W	V	G	V	H	I	Q	Z	H	V	D	L	F	
	3	2	0	3	3	8	4	2	2	2	2	3	7	0	0	3	2	4	4	3	57
1	G	W	R	V	Q	S	M	X	W	H	W	I	J	R	A	I	W	E	M	G	
	2	2	8	2	0	6	2	0	2	3	2	7	0	8	7	7	2	13	2	2	77
2	H	X	S	W	R	T	N	Y	X	I	X	J	K	S	B	J	X	F	N	H	
	3	0	6	2	8	9	8	2	0	7	0	0	0	6	1	0	0	3	8	3	66
3	I	Y	T	X	S	U	O	Z	Y	J	Y	K	L	T	C	K	Y	G	O	I	
	7	2	9	0	6	3	8	0	2	0	2	0	4	9	3	0	2	2	8	7	74
4	J	Z	U	Y	T	V	P	A	Z	K	Z	L	M	U	D	L	Z	H	P	J	
	0	0	3	2	9	2	3	7	0	0	0	4	2	3	4	4	0	3	3	0	49
5	K	A	V	Z	U	W	Q	B	A	L	A	M	N	V	E	M	A	I	Q	K	
	0	7	2	0	3	2	0	1	7	4	7	2	8	2	13	2	7	7	0	0	74
6	L	B	W	A	V	X	R	C	B	M	B	N	O	W	F	N	B	J	R	L	
	4	1	2	7	2	0	8	3	1	2	1	8	8	2	3	8	1	0	8	4	73
7	M	C	X	B	W	Y	S	D	C	N	C	O	P	X	G	O	C	K	S	M	
	2	3	0	1	2	2	6	4	3	8	3	8	3	0	2	8	3	0	6	2	66
8	N	D	Y	C	X	Z	T	E	D	O	D	P	Q	Y	H	P	D	L	T	N	
	8	4	2	3	0	0	9	13	4	8	4	3	0	2	3	3	4	4	9	8	91
9	O	E	Z	D	Y	A	U	F	E	P	E	Q	R	Z	I	Q	E	M	U	O	
	8	13	0	4	2	7	3	3	13	3	13	0	8	0	7	0	13	2	3	8	110
10	P	F	A	E	Z	B	V	G	F	Q	F	R	S	A	J	R	F	N	V	P	
	3	3	7	13	0	1	2	2	3	0	3	8	6	7	0	8	3	8	2	3	82
11	Q	G	B	F	A	C	W	H	G	R	G	S	T	B	K	S	G	O	W	Q	
	0	2	1	3	7	3	2	3	2	8	2	6	9	1	0	6	2	8	2	0	67
12	R	H	C	G	B	D	X	I	H	S	H	T	U	C	L	T	H	P	X	R	
	8	3	3	2	1	4	0	7	3	6	3	9	3	3	4	9	3	3	0	8	82
13	S	I	D	H	C	E	Y	J	I	T	I	U	V	D	M	U	I	Q	Y	S	
	6	7	4	3	3	13	2	0	7	9	7	3	2	4	2	3	7	0	2	6	90
14	T	J	E	I	D	F	Z	K	J	U	J	V	W	E	N	V	J	R	Z	T	
	9	0	13	7	4	3	0	0	0	3	0	2	2	13	8	2	0	8	0	9	83
15	U	K	F	J	E	G	A	L	K	V	K	W	X	F	O	W	K	S	A	U	
	3	0	3	0	13	2	7	4	0	2	0	2	0	3	8	2	0	6	7	3	65
16	V	L	G	K	F	H	B	M	L	W	L	X	Y	G	P	X	L	T	B	V	
	2	4	2	0	3	3	1	2	4	2	4	0	2	2	3	0	4	9	1	2	50
17	W	M	H	L	G	I	C	N	M	X	M	Y	Z	H	Q	Y	M	U	C	W	
	2	2	3	4	2	7	3	8	2	0	2	2	0	3	0	2	2	3	3	2	52
18	X	N	I	M	H	J	D	O	N	Y	N	Z	A	I	R	Z	N	V	D	X	
	0	8	7	2	3	0	4	8	8	2	8	0	7	7	8	0	8	2	4	0	86
19	Y	O	J	N	I	K	E	P	O	Z	O	A	B	J	S	A	O	W	E	Y	
	2	8	0	8	7	0	13	3	8	0	8	7	1	0	6	7	8	2	13	2	103
20	Z	P	K	O	J	L	F	Q	P	A	P	B	C	K	T	B	P	X	F	Z	
	0	3	0	8	0	4	3	0	3	7	3	1	3	0	9	1	3	0	3	0	51
21	A	Q	L	P	K	M	G	R	Q	B	Q	C	D	L	U	C	Q	Y	G	A	
	7	0	4	3	0	2	2	8	0	1	0	3	4	4	3	3	0	2	2	7	55
22	B	R	M	Q	L	N	H	S	R	C	R	D	E	M	V	D	R	Z	H	B	
	1	8	2	0	4	8	3	6	8	3	8	4	13	2	2	4	8	0	3	1	88
23	C	S	N	R	M	O	I	T	S	D	S	E	F	N	W	E	S	A	I	C	
	3	6	8	8	2	8	7	9	6	4	6	13	3	8	2	13	6	7	7	3	129
24	D	T	O	S	N	P	J	U	T	E	T	F	G	O	X	F	T	B	J	D	
	4	9	8	6	8	3	0	3	9	13	9	3	2	8	0	3	9	1	0	4	102
25	E	U	P	T	O	Q	K	V	U	F	U	G	H	P	Y	G	U	C	K	E	
	13	3	3	9	8	0	0	2	3	3	3	2	3	3	2	2	3	3	0	13	78

COLUMN 2

Generatrix																					Frequency value
0	N	P	D	N	N	M	U	G	S	H	G	W	Q	E	N	C	N	S	B	Z	
	8	3	4	8	8	2	3	2	6	3	2	2	0	13	8	3	8	6	1	0	90
1	O	Q	E	O	O	N	V	H	T	I	H	X	R	F	O	D	O	T	C	A	
	8	0	13	8	8	8	2	3	9	7	3	0	8	3	8	4	8	9	3	7	119
2	P	R	F	P	P	O	W	I	U	J	I	Y	S	G	P	E	P	U	D	B	
	3	8	3	3	3	8	2	7	3	0	7	2	6	2	3	13	3	3	4	1	84
3	Q	S	G	Q	Q	P	X	J	V	K	J	Z	T	H	Q	F	Q	V	E	C	
	0	6	2	0	0	3	0	0	2	0	0	0	9	3	0	3	0	2	13	3	46
4	R	T	H	R	R	Q	Y	K	W	L	K	A	U	I	R	G	R	W	F	D	
	8	9	3	8	8	0	2	0	2	4	0	7	3	7	8	2	8	2	3	4	88
5	S	U	I	S	S	R	Z	L	X	M	L	B	V	J	S	H	S	X	G	E	
	6	3	7	6	6	8	0	4	0	2	4	1	2	0	6	3	6	0	2	13	79
6	T	V	J	T	T	S	A	M	Y	N	M	C	W	K	T	I	T	Y	H	F	
	9	2	0	9	9	6	7	2	2	8	2	3	2	0	9	7	9	2	3	3	94
7	U	W	K	U	U	T	B	N	Z	O	N	D	X	L	U	J	U	Z	I	G	
	3	2	0	3	3	9	1	8	0	8	8	4	0	4	3	0	3	0	7	2	68
8	V	X	L	V	V	U	C	O	A	P	O	E	Y	M	V	K	V	A	J	H	
	2	0	4	2	2	3	3	8	7	3	8	13	2	2	2	0	2	7	0	3	73
9	W	Y	M	W	W	V	D	P	B	Q	P	F	Z	N	W	L	W	B	K	I	
	2	2	2	2	2	2	4	3	1	0	3	3	0	8	2	4	2	1	0	7	50
10	X	Z	N	X	X	W	E	Q	C	R	Q	G	A	O	X	M	X	C	L	J	
	0	0	8	0	0	2	13	0	3	8	0	2	7	8	0	2	0	3	4	0	60
11	Y	A	O	Y	Y	X	F	R	D	S	R	H	B	P	Y	N	Y	D	M	K	
	2	7	8	2	2	0	3	8	4	6	8	3	1	3	2	8	2	4	2	0	75
12	Z	B	P	Z	Z	Y	G	S	E	T	S	I	C	Q	Z	O	Z	E	N	L	
	0	1	3	0	0	2	2	6	13	9	6	7	3	0	0	8	0	13	8	4	85
13	A	C	Q	A	A	Z	H	T	F	U	T	J	D	R	A	P	A	F	O	M	
	7	3	0	7	7	0	3	9	3	3	9	0	4	8	7	3	7	3	8	2	93
14	B	D	R	B	B	A	I	U	G	V	U	K	E	S	B	Q	B	G	P	N	
	1	4	8	1	1	7	7	3	2	2	3	0	13	6	1	0	1	2	3	8	73
15	C	E	S	C	C	B	J	V	H	W	V	L	F	T	C	R	C	H	Q	O	
	3	13	6	3	3	1	0	2	3	2	2	4	3	9	3	8	3	3	0	8	79
16	D	F	T	D	D	C	K	W	I	X	W	M	G	U	D	S	D	I	R	P	
	4	3	9	4	4	3	0	2	7	0	2	2	2	3	4	6	4	7	8	3	77
17	E	G	U	E	E	D	L	X	J	Y	X	N	H	V	E	T	E	J	S	Q	
	13	2	3	13	13	4	4	0	0	2	0	8	3	2	13	9	13	0	6	0	108
18	F	H	V	F	F	E	M	Y	K	Z	Y	O	I	W	F	U	F	K	T	R	
	3	3	2	3	3	13	2	2	0	0	2	8	7	2	3	3	3	0	9	8	76
19	G	I	W	G	G	F	N	Z	L	A	Z	P	J	X	G	V	G	L	U	S	
	2	7	2	2	2	3	8	0	4	7	0	3	0	0	2	2	2	4	3	6	59
20	H	J	X	H	H	G	O	A	M	B	A	Q	K	Y	H	W	H	M	V	T	
	3	0	0	3	3	2	8	7	2	1	7	0	0	2	3	2	3	2	2	9	59
21	I	K	Y	I	I	H	P	B	N	C	B	R	L	Z	I	X	I	N	W	U	
	7	0	2	7	7	3	3	1	8	3	1	8	4	0	7	0	7	8	2	3	81
22	J	L	Z	J	J	I	Q	C	O	D	C	S	M	A	J	Y	J	O	X	V	
	0	4	0	0	0	7	0	3	8	4	3	6	2	7	0	2	0	8	0	2	56
23	K	M	A	K	K	J	R	D	P	E	D	T	N	B	K	Z	K	P	Y	W	
	0	2	7	0	0	0	8	4	3	13	4	9	8	1	0	0	0	3	2	2	66
24	L	N	B	L	L	K	S	E	Q	F	E	U	O	C	L	A	L	Q	Z	X	
	4	8	1	4	4	0	6	13	0	3	13	3	8	3	4	7	4	0	0	0	85
25	M	O	C	M	M	L	T	F	R	G	F	V	P	D	M	B	M	R	A	Y	
	2	8	3	2	2	4	9	3	8	2	3	2	3	4	2	1	2	8	7	2	77

FIGURE 12.

d. It will be noted that the frequency value of the 23d generatrix for the first column of cipher letters is the greatest; that of the first generatrix for the second column is the greatest. In both cases these are the correct generatrices. Thus the selection of the correct generatrices in such cases has been reduced to a purely mathematical basis which is at times of much assistance in effecting a quick solution. Moreover, an understanding of the principles involved will be of considerable value in subsequent work.

25. Solution by the probable-word method.—*a.* Occasionally one may encounter a cryptogram which is so short that it contains no recurrences even of digraphs, and thus gives no indications of the number of alphabets involved. If the sliding mixed component is known, one may apply the method illustrated in Par. 15, assuming the presence of a probable word, checking it against the text and the sliding components to establish a key, if the correspondents are using key words.

b. For example, suppose that the presence of the word ENEMY is assumed in the message in Par. 23*b* above. One proceeds to check it against an unknown key word, sliding the already reconstructed mixed component against the normal and starting with the first letter of the cryptogram, in this manner:

When ENEMY equals
SFDZR then A_p successively equals
XENFW

The sequence XENFW spells no intelligible word. Therefore, the location of the assumed word ENEMY is shifted one letter forward in the cipher text, and the test is made again, just as was explained in Par. 15. When the group AQRLU is tried, the key letters ZIMUT are obtained, which, taken as a part of a word, suggests the word AZIMUTH. The method must yield solution when the correct assumptions are made.

c. The danger to cryptographic security resulting from the inclusion of *cryptographed* addresses and signatures in cryptographic messages becomes quite obvious in the light of solution by the probable-word method. To illustrate, reference is made to the message employed in Pars. 19–22. It will be noted in Par. 22*b* that the message carried a signature (Treer, Col.) and that the latter was enciphered. Suppose that this were an authorized practice, and that every message could be assumed to conclude with a cryptographed signature. The signature "TREER COL" would at once afford a very good basis for the quick solution of subsequent messages emanating from the same headquarters as did the first message, because presumably this same signature would appear in other messages. It is for this reason that addresses and signatures must not be cryptographed; if they must be included they should be cryptographed in a totally different system or by a wholly different method, perhaps by means of a special address and signature code. It would be best, however, to omit all addresses and signatures, and to let the call signs of the headquarters concerned also convey these parts of the message, leaving the delivery to the addressee a matter for local action.

26. Solution when the plain component is a mixed sequence, the cipher component, the normal.—*a.* This falls under Case B (2) outlined in Par. 6. It is not the usual method of employing a single mixed component, but may be encountered occasionally in cipher devices.

b. The preliminary steps, as regards factoring to determine the length of the period, are the same as usual. The message is then transcribed into its periods. Frequency distributions are then made, as usual, and these are attacked by the principles of frequency and recurrence. An attempt is made to apply the principles of direct symmetry of position, but this attempt will be futile, for the reason that the plain component is in this case an *unknown* mixed sequence.

(See Par. 18*d*.) Any attempt to find symmetry in the secondary alphabets based upon the normal sequence can therefore disclose no symmetry because the symmetry which exists is based upon a wholly different sequence.

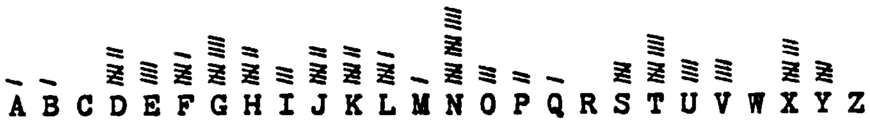
c. However, if the principles of direct symmetry of position are of no avail in this case, there are certain other principles of symmetry which may be employed to great advantage. To explain them an actual example will be used. Let it be assumed that it is known to the cryptanalyst that the enemy is using the general system under discussion, *viz*, a mixed sequence, variable from day to day, is used as plain component; the normal sequence is used as cipher component; and a repeating key, variable from message to message, is used in the ordinary manner.

The following message has been intercepted:

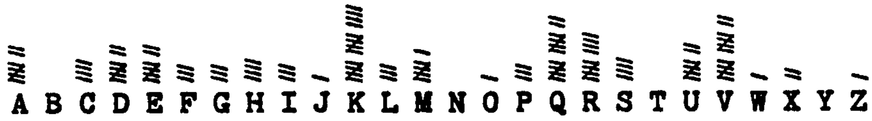
	1	2	3	4	5	6
A.	Q E O V K	L R M L Z	J V G T G	N D L V K	E V N T Y	E R M U E
B.	V R Z M O	Y A A M P	D K E I J	S F M Y O	Y H M M E	G Q A M B
C.	U Q A X R	H U F B U	K Q Y M U	N E L V T	K Q I L E	K Z B U E
D.	U L I B K	N D A X B	X U D G L	L A D V K	P O A Y O	D K K Y K
E.	L A D H Y	B V N F V	U E E M E	F F M T E	G V W B Y	T V D Z L
F.	S P B H B	X V A Z C	U D Y U E	L K M M A	E U D D K	N C F S H
G.	H S A H Y	T M G U J	H Q X P P	D K O U E	X U Q V B	F V W B X
H.	N X A L B	T C D L M	I V A A A	N S Z I L	O V W V P	Y A G Z L
J.	S H M M E	G Q D H O	Y H I V P	N C R R E	X K D Q Z	G K N C G
K.	N Q G U Y	J I W Y Y	T M A H W	X R L B L	O A D L G	N Q G U Y
L.	J U U G B	J H R V X	E R F L E	G W G U O	X E D T P	D K E I Z
M.	V X N W A	F A A N E	M K G H B	S S N L O	K J C B Z	T G G L O
N.	P K M B X	H G E R Y	T M W L Z	N Q C Y Y	T M W I P	D K A T E
P.	F L N U J	N D T V X	J R Z T L	O P A H C	D F Z Y Y	D E Y C L
Q.	G P G T Y	T E C X B	H Q E B R	K V W M U	N I N G J	I Q D L P
R.	J K A T E	G U W B R	H U Q W M	V R Q B W	Y R F B F	K M W M B
S.	T M U L Z	L A A H Y	J G D V K	L K R R E	X K N A O	N D S B X
T.	X C G Z A	H D G T L	V K M B W	I S A U E	F D N W P	N L Z I J
V.	S R Q Z L	A V N H L	G V W V K	F I G H P	G E C Z U	K Q A P

d. A study of the recurrences and factoring their intervals discloses that five alphabets are involved. Uniliteral frequency distributions are made and are shown in Fig. 19*a*:

ALPHABET 1



ALPHABET 2



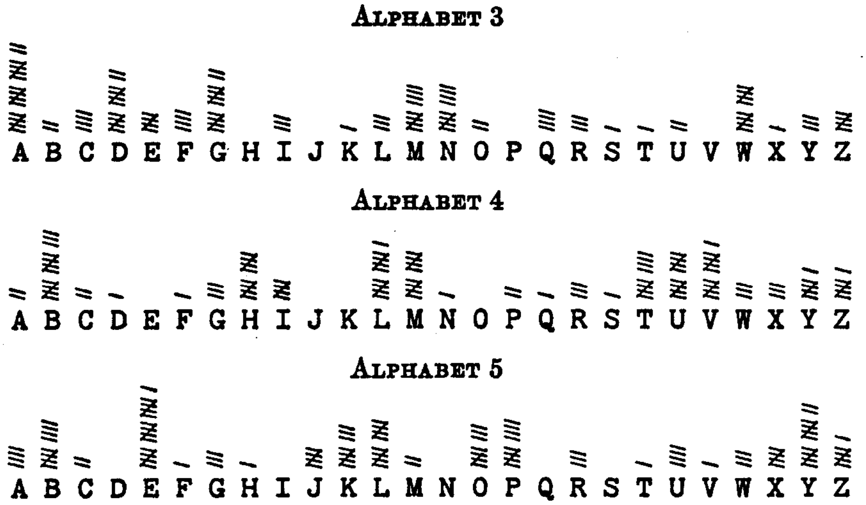


FIGURE 19a.

e. Since the cipher component in this case is the normal alphabet, *it follows that the five frequency distributions are based upon a sequence which is known, and therefore, the five frequency distributions should manifest a direct symmetry of distribution of crests and troughs.* By virtue of this symmetry and by shifting the five distributions relative to one another to proper superimpositions, the several distributions may be combined into a single uniliteral distribution. Note how this shifting has been done in the case of the five illustrative distributions:

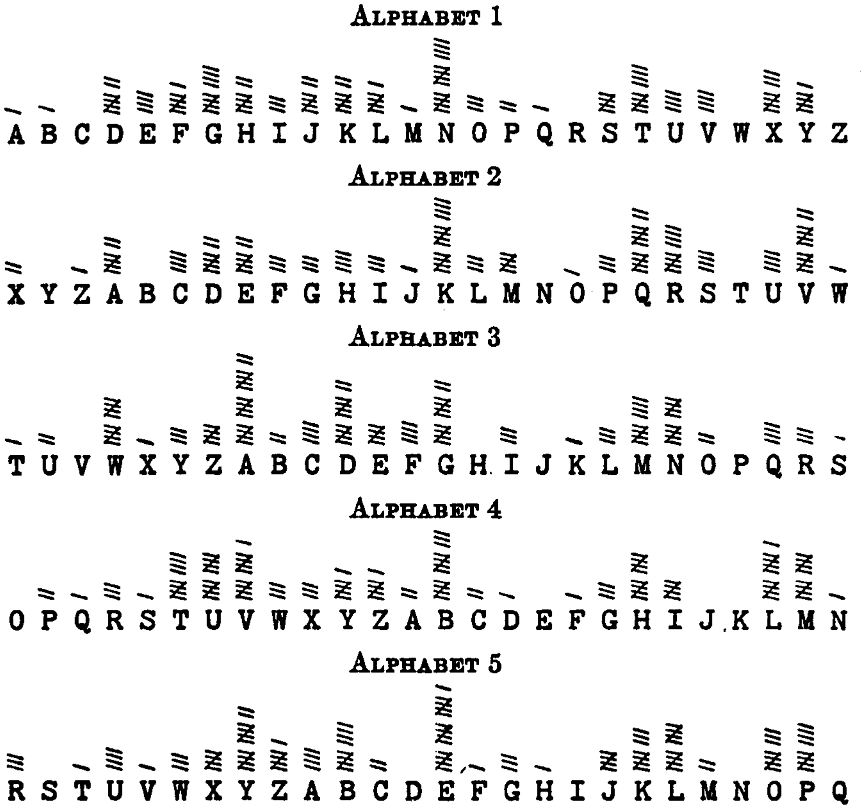


FIGURE 19b.

f. The superimposition of the respective distributions enables one to convert the cipher letters of the five alphabets into one alphabet. Suppose it is decided to convert Alphabets 2, 3, 4, and 5 into Alphabet 1. It is merely necessary to substitute for the respective letters in the four alphabets those which stand above them in Alphabet 1. For example, in Fig. 19*b*, X_e in Alphabet 2 is directly under A_e in Alphabet 1; hence, if the superimposition is correct then $\overset{2}{X_e}=\overset{1}{A_e}$. Therefore, in the cryptogram it is merely necessary to replace every X_e in the second position by A_e. Again T_e in Alphabet 3=A_e in Alphabet 1; therefore, in the cryptogram one replaces every T_e in the third position by A_e. The entire process, hereinafter designated as *conversion into monoalphabetic terms*, gives the following *converted* message:

	1	2	3	4	5	6
A.	Q H V H T	L U T X I	J Y N F P	N G S H T	E Y U F H	E U T G N
B.	V U G Y X	Y D H Y Y	D N L U S	S I T K X	Y K T Y N	G T H Y K
C.	U T H J A	H X M N D	K T F Y D	N H S H C	K T P X N	K C I G N
D.	U O P N T	N G H J K	X X K S U	L D K H T	P R H K X	D N R K T
E.	L D K T H	B Y U R E	U H L Y N	F I T F N	G Y D N H	T Y K L U
F.	S S I T K	X Y H L L	U G F G N	L N T Y J	E X K P T	N F M E Q
G.	H V H T H	T P N G S	H T E B Y	D N V G N	X X X H K	F Y D N G
H.	N A H X K	T F K X V	I Y H M J	N V G U U	O Y D H Y	Y D N L U
J.	S K T Y N	G T K T X	Y K P H Y	N F Y D N	X N K C I	G N U O P
K.	N T N G H	J L D K H	T P H T F	X U S N U	O D K X P	N T N G H
L.	J X B S K	J K Y H G	E U M X N	G Z N G X	X H K F Y	D N L U I
M.	V A U I J	F D H Z N	M N N T K	S V U X X	K M J N I	T J N X X
N.	P N T N G	H J L D H	T P D X I	N T J K H	T P D U Y	D N H F N
P.	F O U G S	N G A H G	J U G F U	O S H T L	D I G K H	D H F O U
Q.	G S N F H	T H J J K	H T L N A	K Y D Y D	N L U S S	I T K X Y
R.	J N H F N	G X D N A	H X X I V	V U X N F	Y U M N O	K P D Y K
S.	T P B X I	L D H T H	J J K H T	L N Y D N	X N U M X	N G Z N G
T.	X F N L J	H G N F U	V N T N F	I V H G N	F G U I Y	N O G U S
V.	S U X L U	A Y U T U	G Y D H T	F L N T Y	G H J L D	K T H B

The uniliteral frequency distribution for this converted text follows. Note that the frequency of each letter is the sum of the five frequencies in the corresponding columns of Fig. 19*b*.

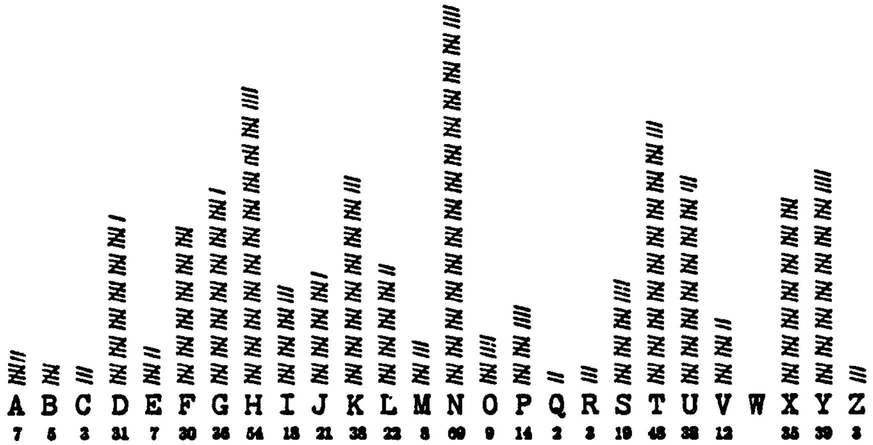


FIGURE 20.

g. The problem having been reduced to monoalphabetic terms, a triliteral frequency distribution can now be made and solution readily attained by simple principles. It yields the following:

JAPAN CONSULTED GERMANY TODAY ON REPORTS THAT THE COMMUNIST INTERNATIONAL WAS BEHIND THE AMAZING SEIZURE OF GENERALISSIMO CHIANG KAI SHEK IN CHINA. TOKYO ACTED UNDER THE ANTICOMMUNIST ACCORD RECENTLY SIGNED BY JAPAN AND GERMANY. THE PRESS SAID THERE WAS INDISPUTABLE PROOF THAT THE COMINTERN INSTIGATED THE SEIZURE OF GENERAL CHIANG AND SOME OF HIS GENERALS. MILITARY OBSERVERS SAID THE COUP WOULD HAVE BEEN IMPOSSIBLE UNLESS GENERAL CHANG HSUEN LIANG HOTHEADED FORMER WAR LORD OF MANCHURIA HAD FORMED AN ALLIANCE WITH THE COMMUNIST LEADERS HE WAS SUPPOSED TO BE FIGHTING. SUCH AN ALLIANCE THESE OBSERVERS DECLARED OPENED UP A RED ROUTE FROM MOSCOW TO NORTH AND CENTRAL CHINA.

h. The reconstruction of the plain component is now a very simple matter. It is found to be as follows:

H Y D R A U L I C B E F G J K M N O P Q S T V W X Z

Note also, in Fig. 19*b*, the keyword for the message, (HEAVY), the letters being in the columns headed by the letter H.

i. The solution of subsequent messages with different keys can now be reached directly, by a simple modification of the principles explained in Par. 18. This modification consists in using for the completion sequence the *mixed plain component* (now known) instead of the normal alphabet, after the cipher letters have been converted into their plain-component equivalents. Let the student confirm this by experiment.

j. The probable-word method of solution discussed under Paragraph 20 is also applicable here, in case of very short cryptograms. This method presupposes of course, possession of the mixed component and the procedure is essentially the same as that in Par. 20. In the example discussed in the present paragraph, the letter A on the plain component was successively set against the key letters HEAVY; but this is not the only possible procedure.

k. The student should go over carefully the principle of "conversion into monoalphabetic terms" explained in subparagraph *f* above until he thoroughly understands it. Later on he will encounter cases in which this principle is of very great assistance in the cryptanalysis of more complex problems. (Another example will be found under Par. 45.)

l. The principle illustrated in subparagraph *e*, that is, shifting two or more monoalphabetic frequency distributions relatively so as to bring them into proper alignment for amalgamation into a single monoalphabetic distribution, is called *matching*. It is a very important cryptanalytic principle. Note that its practical application consists in sliding one monoalphabetic distribution against the other so as to obtain the best coincidence between the *entire sequence* of crests and troughs of one distribution and the *entire sequence* of crests and troughs of the other distribution. When the best point of coincidence has been found, the two sequences may be amalgamated and *theoretically* the single resultant distribution will also be monoalphabetic in character. The successful application of the principle of matching depends upon several factors. First, the cryptographic situation must be such that matching is a correct cryptographic step. For example, the distributions in figure 19*b* are properly subject to matching because the cipher component in the basic sequences concerned in this problem is the normal sequence, while the plain component is a mixed sequence. But it would be futile to try to match the distributions in figure 9, for in that case the cipher component is a mixed sequence, the plain component is the normal sequence. Hence, no amount of shifting or matching can bring the distributions of figure 9 into proper superimposition for correct amalgamation. (If the occurrences in the various distributions in figure 9 had been distributed according to the sequence of letters in the mixed component, then matching would be possible; but in order to be able to distribute these occurrences according to the mixed component, the latter has to be *known*—and that is just what is unknown until the problem has been solved.) A second factor involved in successful matching is the number of elements in the two distributions forming the subject of the test. If both of them have very few tallies, there is hardly sufficient information to permit of matching with any degree of assurance that the work is not in vain. If one of them has many tallies, the other only a few, the chances for success are better than before, because the positions of the *blanks* in the two distributions can be used as a guide for their proper superimposition.

m. There are certain mathematical and statistical procedures which can be brought to bear upon the matter of cryptanalytic matching. These will be presented in a later text. However, until the student has studied these mathematical and statistical methods of matching distributions, he will have to rely upon mere ocular examination as a guide to proper superimposition. Obviously, the more data he has in each distribution, the easier is the correct superimposition ascertained by any method.

SECTION VI

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, II

	Paragraph
Further cases to be considered	27
Identical primary mixed components proceeding in the same direction	28
Cryptographing and decryptographing by means of identical primary mixed components	29
Principles of solution	30

27. **Further cases to be considered.**—*a.* Thus far Cases B (1) and (2), mentioned in Paragraph 6 have been treated. There remains Case B (3), and this case has been further subdivided as follows:

CASE B (3). Both components are mixed sequences.
- (a) Components are identical mixed sequences.
  - (1) Sequences proceed in the same direction (The secondary alphabets are mixed alphabets.)
  - (2) Sequences proceed in opposite directions. (The secondary alphabets are reciprocal mixed alphabets.)
- (b) Components are different mixed sequences. (The secondary alphabets are mixed alphabets.)

b. The first of the foregoing subcases will now be examined.

28. **Identical primary mixed components proceeding in the same direction.**—*a.* It is often the case that the mixed components are derived from an easily remembered word or phrase, so that they can be reproduced at any time from memory. Thus, for example, given the key word QUESTIONABLY, the following mixed sequence is derived:

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z

b. By using this sequence as both plain and cipher component, that is, by sliding this sequence against itself, a series of 26 secondary mixed alphabets may be produced. In enciphering a message, sliding strips may be employed with a key word to designate the particular and successive positions in which the strips are to be set, the same as was the case in previous examples of the use of sliding components. The method of designating the positions, however, requires a word or two of comment at this point. In the examples thus far shown, the key letter, as located on the cipher component, was always set opposite A, as located on the plain component; possibly an erroneous impression has been created, *viz,* that this is invariably the rule. This is decidedly not true, as has already been explained in paragraph 7*c*. If it has seemed to be the case that θ_k always equals A_p, it is only because the text has dealt thus far principally with cases in which the plain component is the normal sequence and its initial letter, which usually constitutes the index for juxtaposing cipher components, is A. It must be emphasized, however, that various conventions may be adopted in this respect; but the most common of them is to employ the initial letter of the plain component as the index letter. That is, the index letter, θ_i, will be the initial letter of the mixed sequence, in this case, Q. Furthermore, to prevent the possibility of ambiguity it will be stated again that the pair of enciphering equations employed in the ensuing discussion will be the first of the 12 set forth under Par. 7*f*, *viz*, $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/1}=\theta_{c/2}$. In this case the subscript "1" means the plain component, the subscript "2", the cipher component, so that the enciphering equation is the following: $\theta_{k/c}=\theta_{i/p}$; $\theta_{p/p}=\theta_{c/c}$.

c. By setting the two sliding components against each other in the two positions shown below, the cipher alphabets labeled (1) and (2) given by two key letters, A and B, are seen to be different.

```
    KEY LETTER=A                                 θ1
                                                 ↓
Plain component....... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
Cipher component......                   QUESTIONABLYCDFGHJKMPRVWXZ
                                                 ↑
                                                 θk
```

Secondary alphabet (1):

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	H	J	P	R	L	V	W	X	D	Z	Q	K	U	G	F	E	A	S	Y	C	B	T	I	O	M	N

```
    KEY LETTER=B                                 θ1
                                                 ↓
Plain component....... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
Cipher component......                  QUESTIONABLYCDFGHJKMPRVWXZ
                                                 ↑
                                                 θk
```

Secondary alphabet (2):

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	J	K	R	V	Y	W	X	Z	F	Q	U	M	E	H	G	S	B	T	C	D	L	I	O	N	P	A

d. Very frequently a quadricular or square table is employed by the correspondents, instead of sliding strips, but the results are the same. The cipher square based upon the word QUESTIONABLY is shown in Fig. 21. It will be noted that it does nothing more than set forth the successive positions of the two primary sliding components; the top line of the square is the plain component, the successive horizontal lines below it, the cipher component in its various juxtapositions. The usual method of employing such a square (i. e., corresponding to the enciphering equations $\theta_{k/c}=\theta_{1/p}$; $\theta_{p/p}=\theta_{c/c}$) is to take as the cipher equivalent of a plain-text letter that letter which lies at the intersection of the vertical column headed by the plain-text letter and the horizontal row begun by the key letter. For example, the cipher equivalent of E_p with keyletter T is the letter O_c; or E_p $(T_k)=O_c$. The method given in paragraph *b*, for determining the cipher equivalents by means of the two sliding strips yields the same results as does the cipher square.

Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z
U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q
E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U
S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E
T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S
I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T
O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I
N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N
B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A
L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B
Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L
C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y
D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C
F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D
G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G
J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H
K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J
M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K
P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M
R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P
V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V
X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W
Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X

FIGURE 21.

29. Cryptographing and decryptographing by identical primary mixed components.—There is nothing of special interest to be noted in connection with the use either of identical mixed components or of an equivalent quadricular table such as that shown in Fig. 21, in enciphering or deciphering a message. The basic principles are the same as in the case of the sliding of one mixed component against the normal, the displacements of the two components being controlled by changeable key words of varying lengths. The components may be changed at will and so on. All this has been demonstrated adequately enough in *Elementary Military Cryptography*, and *Advanced Military Cryptography*.

30. Principles of solution.—*a*. Basically the principles of solution in the case of a cryptogram enciphered by two identical mixed sliding components are the same as in the preceding case. Primary recourse is had to the principles of frequency and repetition of single letters, digraphs, trigraphs, and polygraphs. Once an entering wedge has been forced into the problem, the subsequent steps may consist merely in continuing along the same lines as before, building up the solution bit by bit.

b. Doubtless the question has already arisen in the student's mind as to whether any principles of symmetery of position can be used to assist in the solution and in the reconstruction of the cipher alphabets in cases of the kind under consideration. This phase of the subject will be taken up in the next section and will be treated in a somewhat detailed manner, because the theory and principles involved are of very wide application in cryptanalytics.

SECTION VII

THEORY OF INDIRECT SYMMETRY OF POSITION IN SECONDARY ALPHABETS[1]

Paragraph
Reconstruction of primary components from secondary alphabets.................................... 31

31. Reconstruction of primary components from secondary alphabets.—*a.* Note the two secondary alphabets (1) and (2) given in paragraph 28*c*. Externally they show no resemblance or symmetry despite the fact that they were produced from the same primary components. Nevertheless, when the matter is studied with care, a symmetry of position is discoverable. Because it is a hidden or latent phenomenon, it may be termed *latent symmetry of position*. However, in previous texts the phenomenon has been designated as an *indirect symmetry of position* and this terminology has grown into usage, so that a change is perhaps now inadvisable. Indirect symmetry of position is a very interesting and exceedingly useful phenomenon in cryptanalytics.

b. Consider the following secondary alphabet (the one labeled (2) in paragraph 28*c*):

(2)	Plain......	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher......	J	K	R	V	Y	W	X	Z	F	Q	U	M	E	H	G	S	B	T	C	D	L	I	O	N	P	A

c. Assuming it to be known that this is a secondary alphabet produced by two primary identical mixed components, it is desired to reconstruct the latter. Construct a chain of alternating plain-text and cipher-text equivalents, beginning at any point and continuing until the chain has been completed. Thus, for example, beginning with $A_p=J_c$, $J_p=Q_c$, $Q_p=B_c$, . . ., and dropping out the letters common to successive pairs, there results the sequence A J Q B . . .. By completing the chain the following sequence of letters is established:

A J Q B K U L M E Y P S C R T D V I F W O G X N H Z

d. This sequence consists of 26 letters. *When slid against itself it will produce exactly the same secondary alphabets as do the primary components based upon the word* QUESTIONABLY. To demonstrate that this is the case, compare the secondary alphabets given by the two settings of the externally different components shown below:

Plain component...... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
Cipher component..... QUESTIONABLYCDFGHJKMPRVWXZ

Secondary alphabet (1):

Plain..........	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.........	J	K	R	V	Y	W	X	Z	F	Q	U	M	E	H	G	S	B	T	C	D	L	I	O	N	P	A

Plain component...... AJQBKULMEYPSCRTDVIFWOGXNHZAJQBKULMEYPSCRTDVIFWOGXNHZ
Cipher component..... AJQBKULMEYPSCRTDVIFWOGXNHZ

Secondary alphabet (2):

Plain..........	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.........	J	K	R	V	Y	W	X	Z	F	Q	U	M	E	H	G	S	B	T	C	D	L	I	O	N	P	A

---

[1] After the student has read this and the next section it would be well for him to study Appendix 3, where another and perhaps simpler method is explained.

e. Since the sequence A J Q B K . . . gives exactly the same equivalents in the secondary alphabets as the sequence Q U E S T . . . gives, the former sequence is cryptographically equivalent to the latter sequence. For this reason the A J Q B K . . . sequence is termed an *equivalent primary component*.[1] If the real or original primary component is a key-word mixed sequence, it is hidden or *latent* within the equivalent primary sequence; but it can be made *patent* by decimation of the equivalent primary component. The procedure is as follows: Find three letters in the equivalent primary component such as are likely to have formed an unbroken sequence in the original primary component, and see if the interval between the first and second is the same as that between the second and third. Such a case is presented by the letters W, X, and Z in the equivalent primary component above. Note the sequence . . . W O G X N H Z . . . ; the distance or interval between the letters W, X, and Z is two letters. Continuing the chain by adding letters two intervals removed, the latent original primary component is made patent. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V

f. It is possible to perform the steps given in *c* and *e* in a combined single operation when the original primary component is a key-word mixed sequence. Starting with any pair of letters (in the cipher component of the secondary alphabet) likely to be sequent in the key-word mixed sequence, such as JK, in the secondary alphabet labeled (2), the following chain of digraphs may be set up. Thus, J, K, in the plain component stand over Q, U, respectively, in the cipher component; Q, U, in the plain component stand over B, L, respectively, in the cipher component, and so on. Connecting the pairs in a series, the following results are obtained:

JK → QU → BL → KM → UE → LY → MP → ES → YC → PR → ST → CD → RV →
TI → DF → VW → IO → FG → WX → ON → GH → XZ → NA → HJ → ZQ → AB → JK . . .

These may now be united by means of their common letters:

JK → KM → MP → PR → RV → etc.=J K M P R V W X Z Q U E S T I O N A B L Y C D F G H

The original primary component is thus completely reconstructed.

g. Not all of the 26 secondary alphabets of the series yielded by two sliding primary components may be used to develop a complete equivalent primary component. If examination be made, it will be found that only 13 of these secondary alphabets will yield complete equivalent primary components when the method of reconstruction shown in subparagraph *c* above is followed. For example the following secondary alphabet, which is also derived, from the primary components based upon the word QUESTIONABLY will not yield a complete chain of 26 plain text-cipher-plain text equivalents:

Plain........	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.......	C	D	H	J	O	K	M	P	B	R	V	F	W	Y	L	X	T	Z	N	A	I	Q	U	E	G	S

---

[1] Such an equivalent component is merely a sequence which has been or can be developed or derived from the original sequence or basic primary component by applying a *decimation* process to the latter; conversely, the original or basic component can be derived from an equivalent component by applying the same sort of process to the equivalent component. By decimation is meant the selection of elements from a sequence according to some fixed interval. For example, the sequence A E I M . . . is derived, by decimation, from the normal alphabet by selecting every fourth letter.

Equivalent primary component:

1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	
A	C	H	P	X	E	O	L	F	K	V	Q	T	A	C	H	. . . (The A C H sequence begins again.)

h. It is seen that only 13 letters of the chain have been established before the sequence begins to repeat itself. It is evident that exactly one-half of the chain has been established. The other half may be established by beginning with a letter not in the first half. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	
B	D	J	R	Z	S	N	Y	G	M	W	U	I	B	D	J	. . . (The B D J sequence begins again.)

i. It is now necessary to distribute the letters of each half-sequence within 26 spaces, to correspond with their placements in a complete alphabet. This can only be done by allowing a constant *odd* number of spaces between the letters of one of the half-sequences. Distributions are therefore made upon the basis of 3, 5, 7, 9, . . . spaces. Select that distribution which most nearly coincides with the distribution to be expected in a key-word component. Thus, for example, with the first half-sequence the distribution selected is the one made by leaving three spaces between the letters. It is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	–	L	–	C	–	F	–	H	–	K	–	P	–	V	–	X	–	Q	–	E	–	T	–	O	–

j. Now interpolate, by the same constant interval (three in this case), the letters of the other half-sequence. Noting that the group F – H appears in the foregoing distribution, it is apparent that G of the second half-sequence should be inserted between F and H. The letter which immediately follows G in the second half-sequence, *viz*, M, is next inserted in the position three spaces to the right of G, and so on, until the interpolation has been completed. This yields the original primary component, which is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N

k. Another method of handling cases such as the foregoing is indicated in subparagraph *f* By extending the principles set forth in that subparagraph, one may reconstruct the following chain of 13 pairs from the secondary alphabet given in subparagraph *g:*

1	2	3	4	5	6	7	8	9	10	11	12	13	1
CD →	HJ →	PR →	XZ →	ES →	ON →	LY →	FG →	KM →	VW →	QU →	TI →	AB	→ CD . . .

Now find, in the foregoing chain, two pairs likely to be sequent, for example HJ and KM and count the interval between them in the chain. It is 7 (counting by pairs). If this decimation interval is now applied to the chain of pairs, the following is established:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G

l. The reason why a complete chain of 26 letters cannot be constructed from the secondary alphabet given under subparagraph *g* is that it represents a case in which two primary components of 26 letters were slid an *even* number of intervals apart. (This will be explained in further detail in subparagraph *r* below.) There are in all 12 such cases, none of which will admit of the construction of a complete chain of 26 letters. In addition, there is one case wherein, despite the fact that the primary components are an *odd* number of intervals apart, the secondary alphabet cannot be made to yield a complete chain of 26 letters for an equivalent primary component. This is the case in which the displacement is 13 intervals. Note the secondary alphabet based upon the primary components below (which are the same as those shown in subparagraph *d*):

PRIMARY COMPONENTS

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z
D F G H J K M P R V W X Z Q U E S T I O N A B L Y C

SECONDARY ALPHABET

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	R	V	Z	Q	G	U	E	S	K	T	I	W	O	P	M	N	D	A	H	J	F	B	L	Y	X	C

m. If an attempt is made to construct a chain of letters from this secondary alphabet alone, no progress can be made because the alphabet is completely reciprocal. However, the cryptanalyst need not at all be baffled by this case. The attack will follow along the lines shown below in subparagraphs *n* and *o*.

n. If the original primary component is a key-word mixed sequence, the cryptanalyst may reconstruct it by attempting to "dovetail" the 13 reciprocal pairs (AR, BV, CZ, DQ, EG, FU, HS, IK, JT, LW, MO, NP, and XY) into one sequence. The members of these pairs are all 13 intervals apart. Thus:

```
0  1  2  3  4  5  6  7  8  9  10 11 12 13
A  .  .  .  .  .  .  .  .  .  .  .  .  R
B  .  .  .  .  .  .  .  .  .  .  .  .  V
C  .  .  .  .  .  .  .  .  .  .  .  .  Z
D  .  .  .  .  .  .  .  .  .  .  .  .  Q
E  .  .  .  .  .  .  .  .  .  .  .  .  G
F  .  .  .  .  .  .  .  .  .  .  .  .  U
H  .  .  .  .  .  .  .  .  .  .  .  .  S
I  .  .  .  .  .  .  .  .  .  .  .  .  K
J  .  .  .  .  .  .  .  .  .  .  .  .  T
L  .  .  .  .  .  .  .  .  .  .  .  .  W
M  .  .  .  .  .  .  .  .  .  .  .  .  O
N  .  .  .  .  .  .  .  .  .  .  .  .  P
X  .  .  .  .  .  .  .  .  .  .  .  .  Y
```

FIGURE 22.

Write out the series of numbers from 1 to 26 and insert as many pairs into position as possible. being guided by considerations of probable partial sequences in the key-word mixed sequence, Thus:

```
0  1  2  3  4  5  6  7  8  9  10 11 12 13 14 15 16
A  B  C  D  .  .  .  .  .  .  .  .  .  R  V  Z  Q
```

It begins to look as though the key-word commences with the letter Q, in which case it should be followed by U. This means that the next pair to be inserted is FU. Thus:

```
0  1  2  3  4  5  6  7  8  9  10 11 12 13 14 15 16 17
A  B  C  D  F  .  .  .  .  .  .  .  .  R  V  Z  Q  U
```

The sequence A B C D F means that E is in the key. Perhaps the sequence is A B C D F G H. Upon trial, using the pairs EG and HS, the following placements are obtained:

```
0  1  2  3  4  5  6  7  8  9  10 11 12 13 14 15 16 17 18 19
A  B  C  D  F  G  H  .  .  .  .  .  .  R  V  Z  Q  U  E  S
```

This suggests the word QUEST or QUESTION. The pair JT is added:

```
0  1  2  3  4  5  6  7  8  9  10 11 12 13 14 15 16 17 18 19 20
A  B  C  D  F  G  H  J  .  .  .  .  .  R  V  Z  Q  U  E  S  T
```

The sequence G H J suggests G H J K, which places an I after T. Enough of the process has been shown to make the steps clear.

o. Another method of circumventing the difficulties introduced by the 14th secondary alphabet (displacement interval, 13) is to use it in conjunction with another secondary alphabet which is produced by an even-interval displacement. For example, suppose the following two secondary alphabets are available.[1]

0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	R	V	Z	Q	G	U	E	S	K	T	I	W	O	P	M	N	D	A	H	J	F	B	L	Y	X	C
2	X	Z	E	S	K	T	I	O	R	N	A	Q	B	W	V	L	H	Y	M	P	J	C	D	F	U	G

FIGURE 23.

The first of these secondaries is the 13-interval secondary; the second is one of the even-interval secondaries, from which only half-chain sequences can be constructed. But if the construction be based upon the two sequences, 1 and 2 in the foregoing diagram, the following is obtained:

R X U T N L D H M V Z E I A Y F J P W Q S O B C G K

This is a complete equivalent primary component. The original key-word mixed component can be recovered from it by decimation based upon the 9th interval:

R V W X Z Q U E S T I O N A B L Y C D F G H J K M P

p. (1) When the primary components are identical mixed sequences proceeding in *opposite* directions, all the secondary alphabets will be reciprocal alphabets. Reconstruction of the primary component can be accomplished by the procedure indicated under subparagraph *o* above. Note the following three reciprocal secondary alphabets:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	P	M	H	G	Q	F	D	C	W	Y	L	K	B	R	V	A	E	N	Z	X	U	O	I	T	J	S
2	W	V	M	K	S	J	H	G	Q	F	D	R	C	X	Z	Y	I	L	E	U	T	B	A	N	P	O
3	T	S	Q	Z	L	X	W	V	N	R	P	E	M	I	O	K	C	J	B	A	Y	H	G	F	U	D

FIGURE 24.

(2) Using lines 1 and 2, the following chain can be constructed (equivalent primary component):

P W Q S O B C G K R X U T N L D H M V Z E I A Y F J

---

[1] The method of writing down the secondaries shown in figure 23 will hereafter be followed in all cases when alphabet reconstruction skeletons are necessary. The top line will be understood to be the plain component; it is common to all the secondary alphabets, and is set off from the cipher components by the heavy black line. This top line of letters will be designated by the digit 0, and will be referred to as "the zero line" in the diagram. The successive lines of letters, which occupy the space below the zero line and which contain the various cipher components of the several secondary alphabets, will be numbered serially. These numbers may then be used as reference numbers for designating the horizontal lines in the diagram. The numbers standing above the letters may be used as reference numbers for the vertical columns in the diagram. Hence, any letter in the reconstruction skeleton may be designated by coordinates, giving the horizontal or X coordinate first. Thus, D (2–11) means the letter D standing in line 2, Column 11.

Or, using lines 2 and 3:

W T Y K Z O D P U A G V S L J X I C M Q N F R E B H

The original key-word mixed primary component (based on the word QUESTIONABLY) can be recovered from either of the two foregoing equivalent primary components. But if lines 1 and 3 are used, only half-chains can be constructed:

P T F X A K E C V O H Q L and M S D W N J U Y R I G Z B

This is because 1 and 3 are both odd-interval secondary alphabets, whereas 2 is an even-interval secondary. It may be added that odd-interval secondaries are characterized by having two cases in which a plain-text letter is enciphered by itself; that is, θ_p is identical with θ_c. This phrase "identical with" will be represented by the symbol $\equiv$; the phrase "not identical with" will be represented by the symbol $\not\equiv$. (Note that in secondary alphabet number 1 above, $F_p \equiv F_c$ and $U_p \equiv U_c$; in secondary alphabet number 3 above, $M_p \equiv M_c$ and $O_p \equiv O_c$). This characteristic will enable the cryptanalyst to select at once the proper two secondaries to work with in case several are available; one should show two cases where $\theta_p \equiv \theta_c$; the other should show none.

q. (1) When the primary components are different mixed sequences, their reconstruction from secondary cipher alphabets follows along the same lines as set forth above, under *b* to *j*, inclusive, with the exception that the selection of letters for building up the chain of equivalents for the primary cipher component is restricted to those below the zero line in the reconstruction skeleton. Having reconstructed the primary cipher component, the plain component can be readily reconstructed. This will become clear if the student will study the following example.

0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	T	V	A	B	U	L	I	Q	X	Y	C	W	S	N	D	P	F	E	Z	G	R	H	J	K	M	O
2	Z	J	S	T	V	I	Q	R	M	O	N	K	X	E	A	G	B	W	P	L	H	Y	C	D	F	U

FIGURE 25.

(2) Using only lines 1 and 2, the following chain is constructed:

T Z P G L I Q R H Y O U V J C N E W K D A S X M F B

This is an equivalent primary cipher component. By finding the values of the successive letters of this chain in terms of the plain component of secondary alphabet number 1 (the zero line), the following is obtained:

T Z P G L I Q R H Y O U V J C N E W K D A S X M F B
A S P T F G H U V J Z E B W K N R L X O C M I Y Q D

The sequence A S P T . . . is an equivalent primary plain component. The original key-word mixed components may be recovered from each of the equivalent primary components. That for the primary plain component is based upon the key PUBLISHERS MAGAZINE; that for the primary cipher component is based upon the key QUESTIONABLY.

(3) Another method of accomplishing the process indicated above can be illustrated graphically by the following two chains, based upon the two secondary alphabets set forth in subparagraph *q* (1):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	T	V	A	B	U	L	I	Q	X	Y	C	W	S	N	D	P	F	E	Z	G	R	H	J	K	M	O
2	Z	J	S	T	V	I	Q	R	M	O	N	K	X	E	A	G	B	W	P	L	H	Y	C	D	F	U

Col. 1.		Col. 2.					
A (0–1)	→	T (1–1);	→	T (2–4)	→	D (0–4);	→
D (0–4)	→	B (1–4);	→	B (2–17)	→	Q (0–17);	→
Q (0–17)	→	F (1–17);	→	F (2–25)	→	Y (0–25);	→
Y (0–25)	→	M (1–25);	→	M (2–9)	→	I (0–9);	→
I (0–9)	→	X (1–9);	→	X (2–13)	→	M (0–13);	→
M (0–13)	→	S (1–13);	→	S (2–3)	→	C (0–3);	→
etc.		etc.					

FIGURE 28.

(4) By joining the letters in Column 1, the following chain is obtained: A D Q Y I M, etc. If this be examined, it will be found to be an equivalent primary of the sequence based upon PUBLISHERS MAGAZINE. By joining the letters in Column 2, the following chain is obtained: T B F M X S. This is an equivalent primary of the sequence based upon QUESTIONABLY.

r. A final word concerning the reconstruction of primary components in general may be added. It has been seen that in the case of a 26-element component sliding against itself (both components proceeding in the same direction), it is only the secondary alphabets resulting from odd-interval displacements of the primary components which permit of reconstructing a single 26-letter chain of equivalents. This is true except for the 13th interval displacement, which even though an odd number, still acts like an even number displacement in that no complete chain of equivalents can be established from the secondary alphabet. This exception gives the clue to the basic reason for this phenomenon: it is that the number 26 has two factors, 2 and 13, which enter into the picture. With the exception of displacement-interval 1, *any displacement interval which is a sub-multiple of, or has a factor in common with the number of letters in the primary sequence will yield a secondary alphabet from which no complete chain of 26 equivalents can be derived for the construction of a complete equivalent primary component.* This general rule is applicable only to components which progress in the same direction; if they progress in opposite directions, all the secondary alphabets are reciprocal alphabets and they behave exactly like the reciprocal secondaries resulting from the 13-interval displacement of two 26-letter identical components progressing in the same direction.

s. The foregoing remarks give rise to the following observations based upon the general rule pointed out above. Whether or not a complete equivalent primary component is derivable by decimation from an original primary component (and if not, the lengths and numbers of chains of letters, or incomplete components, that can be constructed in attempts to derive such equivalent components) will depend upon the number of letters in the original primary component and the specific decimation interval selected. For example, in a 26-letter original primary component, decimation interval 5 will yield a complete equivalent primary component of 26 letters, whereas decimation intervals 4 or 8 will yield 2 chains of 13 letters each. In a 24-letter component, decimation interval 5 will also yield a complete equivalent primary component (of 24 letters), but decimation interval 4 will yield 6 chains of 4 letters each, and decimation interval 8 will yield 3 chains of 8 letters each. It also follows that in the case of an original primary component in which the total number of characters is a prime number, *all* decimation intervals will yield complete equivalent primary components. The following table has been drawn up in the light of these observations, for original primary sequences from 16 to 32 elements. (All prime-number sequences have been omitted.) In this table, the column at the extreme left gives the various decimation intervals, omitting in each case the first interval, which merely gives the original primary sequence, and the last interval, which merely gives the original sequence reversed. The top line of the table gives the various lengths of original primary sequences from 32 down to 16. (The student should bear in mind that sequences containing characters in addition to the letters of the alphabet may be encountered; he can add to this table when he is interested in sequences of more than 32 characters.) The numbers within the table then show, for each combination of decimation interval and length of, original sequence, the lengths of the chains of characters that can be constructed. (The student may note the symmetry in each column.) The bottom line shows the total number of complete equivalent primary components which can be derived for each different length of original component.

Decimation interval	Number of characters in original primary component											
	32	30	28	27	26	25	24	22	21	20	18	16
2	16	15	14	27	13	25	12	11	21	10	9	8
3	32	10	28	9	26	25	8	22	7	20	6	16
4	8	15	7	27	13	25	6	11	21	5	9	4
5	32	6	28	27	26	5	24	22	21	4	18	16
6	16	5	14	9	13	25	4	11	7	10	3	8
7	32	30	4	27	26	25	24	22	3	20	18	16
8	4	15	7	27	13	25	6	11	21	5	9	2
9	32	10	28	9	26	25	8	22	7	20	2	16
10	16	3	14	27	13	5	12	11	21	2	9	8
11	32	30	28	27	26	25	24	2	21	20	18	16
12	8	5	7	9	13	25	2	11	7	5	3	4
13	32	30	28	27	2	25	24	22	21	20	18	16
14	16	15	2	27	13	25	12	11	3	10	9	8
15	32	2	28	9	26	5	8	22	7	4	6	
16	2	15	7	27	13	25	6	11	21	5	9	
17	32	30	28	27	26	25	24	22	21	20		
18	16	5	14	9	13	25	4	11	7	10		
19	32	30	28	27	26	25	24	22	21			
20	8	3	7	27	13	5	6	11				
21	32	10	4	9	26	25	8					
22	16	15	14	27	13	25	12					
23	32	30	28	27	26	25						
24	4	5	7	9	13							
25	32	6	28	27								
26	16	15	14									
27	32	10										
28	8	15										
29	32											
30	16											
Total number of complete sequences	14	6	10	16	10	18	16	8	10	6	4	6

SECTION VIII

APPLICATION OF PRINCIPLES OF INDIRECT SYMMETRY OF POSITION

Paragraph
Applying the principles to a specific example........ 32
The cryptogram employed in the exposition........ 33
Fundamental theory........ 34
Application of principles........ 35
General remarks........ 36

32. Applying the principles to a specific example.—*a.* The preceding section, with the many details covered, now forms a sufficient base for proceeding with an exposition of how the principles of indirect symmetry of position can be applied very early in the solution of a polyalphabetic substitution cipher in which sliding primary components were employed to produce the secondary cipher alphabets for the enciphering of the cryptogram.

b. The case described below will serve not only to explain the method of applying these principles but will at the same time show how their application greatly facilitates the solution of a single, rather difficult, polyalphabetic substitution cipher. It is realized, of course, that the cryptogram could be solved by the usual methods of frequency and long, patient experimentation. However, the method to be described was actually applied and very materially reduced the amount of time and labor that would otherwise have been required for solution.

33. The cryptogram employed in the exposition.—*a.* The problem that will be used in this exposition involves an actual cryptogram submitted for solution in connection with a cipher device having two concentric disks upon which the same random mixed alphabet appears, both alphabets progressing in the same direction. This was obtained from a study of the descriptive circular accompanying the cryptogram. By the usual process of factoring, it was determined that the cryptogram involved 10 alphabets. The message as arranged according to its period is shown in Figure 27, in which all repetitions of two or more letters are indicated.

b. The triliteral frequency distributions are given in Figure 28. It will be seen that on account of the brevity of the message, considering the number of alphabets involved, the frequency distributions do not yield many clues. By a very careful study of the repetitions, tentative individual determinations of values of cipher letters, as illustrated in Figures 29, 30, 31, and 32, were made. These are given in sequence and in detail in order to show that there is nothing artificial or arbitrary in the preliminary stages of analysis here set forth.

The Cryptogram

(Repetitions underlined)

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
B	G	B	Z	D	P	F	B	O	U	O
C	G	R	F	T	Z	M	Q	M	A	V
D	K	Z	U	G	D	Y	F	T	R	W
E	G	J	X	N	L	W	Y	O	U	X
F	I	K	W	E	P	Q	Z	O	K	Z
G	P	R	X	D	W	L	Z	I	C	W
H	G	K	Q	H	O	L	O	D	V	M
I	G	O	X	S	N	Z	H	A	S	E
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
N	G	S	X	E	R	O	Z	J	S	E
O	G	V	Q	W	E	J	M	K	G	H

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
S	C	F	B	S	C	V	X	C	H	Q
T	Z	T	Z	S	D	M	X	W	C	M
U	R	K	U	H	E	Q	E	D	G	X
V	F	K	V	H	P	J	J	K	J	Y
W	Y	Q	D	P	C	J	X	L	L	L
X	G	H	X	E	R	O	Q	P	S	E
Y	G	K	B	W	T	L	F	D	U	Z
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
BB	H	S	P	O	P	N	M	D	L	M
CC	G	C	K	W	D	V	B	L	S	E
DD	G	S	U	G	D	P	O	T	H	X

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
FF	L	F	U	Y	D	T	Z	V	H	Q
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
PP	U	K	W	P	E	F	X	E	N	F
QQ	C	C	U	G	D	W	P	E	U	H
RR	Y	B	W	E	W	V	M	D	W	J
SS	R	Z	X							

Figure 27.

TRILITERAL FREQUENCY DISTRIBUTIONS

I

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	EB	FF			XK	YB	ES	XK	ZC	VZ	WQ			ZC	ZR	DC	HC	HR		MK		–F		YQ	QT
	HZ	FC				OR	NH		VQ	ZL	JF						MK							NT	QG
	XK					WJ	ZO		QJ								JZ							NU	
	WG					WK																		HB	
	QK					MO																			
						ES																			
						EV																			
						LH																			
						EK																			
						MC																			
						ES																			

II

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	GZ	QB			WU	ZW	GX		GX	IW	KB			GX		LZ	GF	GX	ZZ	YX	GQ				KU
	BJ	JQ			CB	BB	HV		JU	GQ				HZ		YD	PX	HP	YX						BZ
	YW	RV			LU					RU						JW	SQ	GU							RX
		OD								FV															
		GK								GB															
		CU								BD															
										BX															
										UW															

III

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	CZ		QP		RT				BI	CW					SO	KH				FP	CO	KE	JN		BD
	FS		CH													CR				ZG	KH	GN	RD		QA
	KW		KZ													RE				KH	HL	QO	OS		ZC
	LP															VW				SG		KP	SE		TS
	GW																			FY		BE	HE		OO
																				JG			TC		
																				CG			KD		
																							UO		
																							Z–		

IV

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ZA		ZK	ZP	WP		UD	QO	JP			VA		XL	VP	UC		QQ	XN	FZ			QE		UD	BE
		XD	XW	QW		UD	UE						WK	PP	DC			BC				BT			DF
			XS	XR		UD	VP							WO	BC			ZD				KD			
				XR		UD	DW							XP	WE							BW			
				WW										ZW											

V

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
AA		PF	GY	ZX	ZM					CQ	NW		SZ	HL	DF	RF	EO	DO	WL			DL			TM
LQ		SV	SM	WJ						NX				OT	EQ		EO					EM			
		PJ	WV	HQ											IQ							HM			
		PJ	GP	PF											ON							WO			
			YT												HJ							OM			
			GP												ON							EV			
			GW												OP										
			GW																						

VI

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
AM					CO				EM		WZ	ZQ	PB	RZ	DO	PZ			DZ		CX	LY	EQ	DF	NH
					PB				PJ		OO	WL	PM	RQ	DM	PF			OT		DB	DQ	KJ		
					QV				CX		TF	DX		WQ	PY	KO					WM	DP			
					EX				CO			WZ		SZ		EE									
												FT				AQ									
												WX													

VII

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	FO			QD	YT		ZA		JK		MN	JK		FC	WE	MM			MG		FM		VC	WO	QO
	NL				QJ				XT			AD		LD		XT			TN				MW	PO	LI
	VL				LD							ND		QI		OP							JL		OJ
												PV		JT		OR							MC		MT
												VD		PT		QV							FE		TV
																WR									OR

VIII

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
HS		OJ	OV	XN		TQ		ZC	FH	MG	BC	QA	LA	BU	QS		QG		FR		ZH	XC			
		XH	MC	PU				OK	ZS	JJ	XL	VL	TV	YU			ZS		QX		ML				
		XG	EG								BS			ZK			QV		ZU		QA				
			FU											YX					OX						
			ML																OH						
			MY																JR						

IX

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
MV		IW				KH	JD		CY	OZ	MH		EF			GJ	TW	AE		OO	DM		TZ	DJ	
NE		LW				DX	CQ		KY	IF	LL						TN	JE		OX	NQ		TE		
VV		DH				RN	TX				DM							PE		DZ	RM		OZ		
		WM				CQ	VQ				VW							LE		TZ					
																		RN		EH					

X

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Z	Y	Z
			HQ	SB	KC		LS		QL		LG	VG	RY	UG		HZ					AK	RG	UI	JG	KP
				AG	NC		GR		YR			CR	GH			HZ					AJ	CG	GF	JY	XJ
				SG			CB					LG	SY			VB						CL	HB		UO
				SG			UY					VU				GJ						LB			UK
				XH																					XH
				SG																					

FIGURE 28.

INITIAL VALUES FROM ASSUMPTIONS

$\overset{1}{G}_c=E_p$; $\overset{2}{K}_c=E_p$; $\overset{3}{X}_c=E_p$; and $\overset{5}{D}_c=E_p$, from frequency considerations.

$\overset{3}{U}\overset{4}{G}\overset{5}{D}=THE$; $\overset{4}{P}\overset{5}{C}\overset{6}{J}=THE$; and $\overset{9}{S}\overset{10}{E}\overset{1}{G}=THE$, from study of repetitions.

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
			T	T	H					
B	G	B	Z	D	P	F	B	O	U	O
	E									
C	G	R	F	T	Z	M	Q	M	A	V
	E									
D	K	Z	U	G	D	Y	F	T	R	W
			T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E								
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E								
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
										H
N	G	S	X	E	R	O	Z	J	S	E
	E		E						T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
S	C	F	B	S	C	V	X	C	H	Q
					H					
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E				E				
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E						T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
				T	H	E				H
BB	H	S	P	O	P	N	M	D	L	M
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E		T	H	E					

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
			T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E						T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T						
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
SS	R	Z	X							
			E							

FIGURE 20.

Additional Values from Assumptions (I)

Refer to line DD in Figure 29; $\overset{2}{S}_e$ assumed to be N_p.

Refer to line M in figure 29; $\overset{9}{A}_e$ assumed to be W_p.

Then in lines C–D, $\overset{9}{A}\ \overset{10}{V}\ \overset{1}{K}\ \overset{2}{Z}\ \overset{3}{U}\ \overset{4}{G}\ \overset{5}{D}$ is assumed to be WITH THE.

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
			T	T	H					
B	G	B	Z	D	P	F	B	O	U	O
	E									
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E								
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E								
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
									W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E						T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H								
S	C	F	B	S	C	V	X	C	H	Q
					H					
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E				E				
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E						T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E				
BB	H	S	P	O	P	N	M	D	L	M
		N								
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	N	T	H	E					
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
			T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E						T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T						
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
SS	R	Z	X							
		H	E							

Figure 30.

ADDITIONAL VALUES FROM ASSUMPTIONS (II)

```
                             1 2 3 4 5 6 7 8 9 10
Refer to Figure 30, line A;  W F U P C F O C J Y; assume to be BUT THOUGH.
                             - - T T H - - - - -

                                                  3 4 5 6
Refer to Figure 30, lines N and X, where repetition X E R O occurs; assume EACH
                                                  E - - -
```

```
    1 2 3 4 5 6 7 8 9 10
A   W F U P C F O C J Y
    B U T T H O U G H
B   G B Z D P F B O U O
    E         O
C   G R F I Z M Q M A V
    E               W I
D   K Z U G D Y F T R W
    T H T H E
E   G J X N L W Y O U X
    E   E
F   I K W E P Q Z O K Z
      E   A
G   P R X D W L Z I C W
        E
H   G K Q H O L O D V M
    E E         U
I   G O X S N Z H A S E
    E   E           T H
J   B B J I P O F J H D

K   Q C B Z E X Q T X Z

L   J C Q R Q F V M L M
              O
M   S R Q E W M L N A E
          A         W H
N   G S X E R O Z J S E
    E N E A C H     T H
O   G V Q W E J M K G H
    E         E
```

```
    1 2 3 4 5 6 7 8 9 10
P   R C V O P N B L C W

Q   L Q Z A A A M D C H

R   B Z Z C K Q O I K F
      H         U
S   C F B S C V X C H Q
      U     H     G
T   Z T Z S D M X W C M
            E
U   R K U H E Q E D G X
      E T
V   F K V H P J J K J Y
      E       E E   H
W   Y Q D P C J X L L L
          T H E
X   G H X E R O Q P S E
    E   E A C H     T H
Y   G K B W T L F D U Z
    E E
Z   O C D H W M Z T U Z

AA  K L B F C J O T X E
    T     T H E U     H
BB  H S P O P N M D L M
      N
CC  G C K W D V B L S E
    E       E       T H
DD  G S U G D P O T H X
    E N T H E   U
```

```
    1 2 3 4 5 6 7 8 9 10
EE  B K D Z F M T G Q J
      E
FF  L F U Y D T Z V H Q
      U T   E
GG  Z G W N K X J T R N

HH  Y T X C D P M V L W
        E   E
II  B G B W W O Q R G N
              H
JJ  H H V L A Q Q V A V
                    W I
KK  J Q W O O T T N V Q

LL  B K X D S O Z R S N
      E E     H     T
MM  Y U X O P P Y O X Z

NN  H O Z O W M X C G Q
                  G
OO  J J U G D W Q R V M
        T H E
PP  U K W P E F X E N F
      E   T   O
QQ  C C U G D W P E U H
        T H E
RR  Y B W E W V M D Y J
          A
SS  R Z X
      H E
```

FIGURE 31.

ADDITIONAL VALUES FROM ASSUMPTIONS (III)

$\overset{4\ 5\ 6}{\text{OPN}}$—assume ING from repetition and frequency.

$\overset{9\ 10\ 1}{\text{HQZ}}$—assume ING from repetition and frequency.

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
	B	U	T	T	H	O	U	G	H	
B	G	B	Z	D	P	F	B	O	U	O
	E				N	O				
C	G	R	F	I	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E		A	N					
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E					U			
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
					N				I	
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	O	R	Q	F	V	M	L	M
						O				
M	S	R	Q	E	W	M	L	N	A	E
				A					W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E	A	C	H			T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				
P	R	C	V	O	P	N	B	L	C	W
				I	N	G				
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H					U			
S	C	F	B	S	C	V	X	C	H	Q
		U			H			G	I	N
T	Z	T	Z	S	D	M	X	W	C	M
	G				E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E			N	E	E		H	
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E	A	C	H			T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E	U			H
BB	H	S	P	O	P	N	M	D	L	M
		N		I	N	G				
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	H	T	H	E		U		I	
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
		U	T		E				I	N
GG	Z	G	W	N	K	X	J	T	R	N
	G									
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
						H				
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
				I						N
LL	B	K	X	D	S	O	Z	R	S	N
		E	E			H			T	
MM	Y	U	X	O	P	P	Y	O	X	Z
				I	N					
NN	H	O	Z	O	W	M	X	C	G	Q
				I				G		N
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T		O				
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
				A						
SS	R	Z	X							
		H	E							

FIGURE 32.

c. From the initial and subsequent tentative identifications shown in Figures 29, 30, 31, and 32, the values obtained were arranged in the form of the secondary alphabets in a reconstruction skeleton, shown in Figure 33.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1		W			G		Z													K						
2					K			Z						S							F					
3					X															U						
4	E							G	O											P						
5			R		D			C						P												
6					J		N	O							F											
7																					O					
8							C																			
9								J	H											S			A			
10								E	V					Q												

FIGURE 33.

34. Fundamental theory.—*a.* In paragraph 31, methods of reconstructing primary components from secondary alphabets were given in detail. It is necessary that those methods be fully understood before the following steps be studied. It was there shown that the primary component can be one of a series of equivalent primary sequences, all of which will give exactly similar results so far as the secondary alphabets and the cryptographic text are concerned. It is not necessary that the identical or original primary component employed in the cryptographing be reconstructed; any equivalent primary sequence will serve. The whole question is one of establishing a sequence of letters the interval between which is either identical with that in the original primary component or else is an exact constant multiple of the interval separating the letters in the original primary component. For example, suppose K P X N Q forms a sequence in the original primary component. Here the interval between K and P, and P and X, X and N, N and Q is one; in an equivalent primary component, say the sequence K . . P . . X . . N . . Q, the interval between K and P is three, that between P and X also three, and so on; and the two sequences will yield the same secondary alphabets. So long as the interval between K and P, P and X, X and N, N and Q, . . ., is a constant one, the sequence will be cryptographically equivalent to the original primary sequence and will yield the same secondary alphabets as do those of the original primary sequence. However, in the case of a 26-letter component, it is necessary that this interval be an odd number other than 13, as these are the only cases which will yield one unbroken sequence of 26 letters. Suppose a secondary alphabet to be as follows:

(1)																										
Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher														X		K	N							P		

It can be said that the primary component contains the following sequences:

XN KP NQ PX

These, when united by means of their common letters, yield K P X N Q.

Suppose also the following secondary alphabet is at hand:

(2)	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher														P			X							K		N

Here the sequences PN, XQ, KX, and NZ can be obtained, which when united yield the two sequences KXQ and PNZ.

By a comparison of the sequences K P X N Q, K X Q, and P N Z, one can establish the following:

K	P	X	N	Q	
K	.	X	.	Q	
	P	.	N	.	Z

It follows that one can now add the letter Z to the sequence, making it K P X N Q Z.

b. The reconstruction of a primary component from one of the secondary alphabets by the process given in paragraph 31 requires a complete or nearly complete secondary alphabet. This is at hand only *after* a cryptogram has been completely solved. But if one could employ several very scant or skeletonized secondary alphabets simultaneously with the analysis of the cryptogram, one could then possibly build up a primary component from fewer data and thus solve the cryptogram much more rapidly than would otherwise be possible.

c. Suppose only the cipher components of the two secondary alphabets (1) and (2) given above be placed into juxtaposition. Thus:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
(1)	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.	K	N	.	.	.	.	.	.	P	.	.
(2)	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.	.	X	.	.	.	.	.	.	K	.	N

The sequences PX, XN, and KP are given by juxtaposition. These, when united, yield KPXN as part of the primary sequence. It follows, therefore, that *one can employ the cipher components of secondary alphabets as sources of independent data* to assist in building up the primary sequences. The usefulness of this point will become clearer subsequently.

35. Application of principles.—*a*. Refer now to the reconstruction skeleton shown in Figure 33. Hereafter, in order to avoid all ambiguity and for ease in reference, the position of a letter in Figure 33 will be indicated as stated in footnote 1, page 56. Thus, N (6–7) refers to the letter N in line 6 and in column 7 of Figure 33.

b. (1) Now, consider the following pairs of letters:

E (0–5) J (6–5)
G (0–7) N (6–7)
H (0–8) O (6–8) } HO, OF=HOF
O (0–15) F (6–15)

(One is able to use the line marked zero in Figure 33 since this is a mixed sequence sliding against *itself*.)

(2) The immediate results of this set of values will now be given. Having HOF as a sequence, with EJ as belonging to the same displacement interval, suppose HOF and EJ are placed into juxtaposition as portions of sliding components. Thus:

Plain......	. . . H O F . . .
Cipher....	. . . E J . . . .

When $H_p=E_c$, then $O_p=J_c$.

(3) Refer now to alphabet 10, Figure 33, where it is seen that $H_p=E_c$. *The derived value,* $O_p=J_c$, *can immediately be inserted in the same alphabet* and substituted in the cryptogram.

(4) The student may possibly get a clearer idea of the principles involved if he will regard the matter as though he were dealing with arithmetical proportion. For instance, given any three terms in the proportion 2:8=4:16, the 4th term can easily be found. Furthermore, given the pair of values on the left-hand side of the equation, one may find numerous pairs of values which may be inserted in the right-hand side, or vice versa. For instance, 2:8=4:16 is the same as 2:8=5:20, or 9:36=4:16, and so on. An illustration of each of these principles will now be given, reference being made to Figure 33. As an example of the first principle, note that E (0-5):H (0-8)=J (6-5):O (6-8). Now find E (10-8):H (0-8)=? (10-15):O (0-15). It is clear that J may be inserted as the 3d term in this proportion, thus giving the important new value, $\overset{10}{O_p}=J_c$, which is exactly what was obtained directly above, by means of the partial sliding components. As an example of the second principle, note the following pairs:

E (0-5)	H (0-8)
K (2-5)	Z (2-8)
D (5-5)	C (5-8)
J (6-5)	O (6-8)

These additional pairs are also noted:

K (1-20)	Z (1-7)
T (0-20)	G (0-7)

Therefore, E:H=K:Z=D:C=J:O=T:G, and T may be inserted in position (4-5).

c. (1) Again, GN belongs to the same set of displacement-interval values as do EJ and HOF. Hence, by superimposition:

Plain......	. . . H O F . . .
Cipher....	. . . G N . . . .

(2) Referring to alphabet 4, when $H_p=G_c$, then $\underset{4}{O_p}=N_c$. Therefore, the letter N can be inserted in position (4-15) in Figure 33, and the value $N_c=O_p$ can be substituted in the cryptogram.

(3) Furthermore, note the corroboration found from this particular superimposition:

H (0-8)	G (0-7)
O (6-8)	N (6-7)

This checks up the value in alphabet 6, $G_p=N_c$.

d. (1) Again superimpose HOF and GN:

. . . H O F . . .
. . . . G N . . .

(2) Note this corroboration:

O (6-8)	G (4-8)
F (6-15)	N (4-15)

which has just been inserted in Figure 33, as stated above.

e. (1) Again using HOF and EJ, but in a different superimposition:

. . . H O F . .
. . E J . . . .

(2) Refer now to H (9–9), J (9–8). Directly under these letters is found V (10–9), E (10–8). Therefore, the V can be added immediately before H O F, making the sequence V H O F.

f. (1) Now take V H O F and juxtapose it with E J, thus:

. . . V H O F . . .
. . . E J . . .

(2) Refer now to Figure 33, and find the following:

V (10–9)	E (10–8)
H (9–9)	J (9–8)
O (4–9)	G (4–8)
I (0–9)	H (0–8)

(3) From the value O G it follows that G can be set next to J in E J. Thus:

. . . V H O F . . .
. . . E J G . . .

(4) But G N already is known to belong to the same set of displacement-interval values as E J. Therefore, it is now possible to combine E J, J G, and G N into one sequence, E J G N, yielding:

. . . V H O F . . .
. . . E J G N . . .

g. (1) Refer now to Figure 33.

V (0–22)	E (0–5)
? (1–22)	G (1–5)
? (2–22)	K (2–5)
? (3–22)	X (3–5)
? (5–22)	D (5–5)
? (6–22)	J (6–5)

(2) The only values which can be inserted are:

O (1–22)	G (1–5)
H (6–22)	J (6–5)

(3) This means that $V_p=O_c$ in alphabet 1 and that $V_p=H_c$ in alphabet 6. There is one O_c in the frequency distribution for alphabet 1, and no H_c in that for alphabet 6. The frequency distribution is, therefore, corroborative insofar as these values are concerned.

(*h*) (1) Further, taking E J G N and V H O F, superimpose them thus:

. . . E J G N . . .
. . . . V H O F . . .

(2) Refer now to Figure 33.

E (0–5)	H (0–8)
G (1–5)	? (1–8)

(3) From the diagram of superimposition the value G (1–5) F (1–8) can be inserted, which gives $H_p=F_c$ in alphabet 1.

i. (1) Again, V H O F and E J G N are juxtaposed:

. . . V H O F . . .
. . . E J G N . . .

(2) Refer to Figure 33 and find the following:

H (0–8) G (4–8)
A (0–1) E (4–1)

This means that it is possible to add A, thus:

. . . A V H O F . . .
. . . E J G N . . .

(3) In the set there are also:

E (0–5) G (1–5)
G (0–7) Z (1–7)

Then in the superimposition

. . . E J G N . . .
. . . E J G N . . .

It is possible to add Z under G, making the sequence E J G N Z.

(4) Then taking

. . . A V H O F . . .
. . . E J G N Z . . .

and referring to Figure 33:

H (0–8) N (0–14)
O (6–8) ? (6–14)

It will be seen that O=Z from superimposition, and hence in alphabet 6 $N_p=Z_c$, an important new value, but occurring only once in the cryptogram. Has an error been made? The work so far seems too corroborative in interlocking details to think so.

j. (1) The possibilities of the superimposition and sliding of the AVHOF and the EJGNZ sequences have by no means been exhausted as yet, but a little different trail this time may be advisable.

E (0–5) T (0–20)
G (1–5) K (1–20)
X (3–5) U (3–20)

(2) Then:

. . . E J G N Z . . .
. . . T . K . . .

(3) Now refer to the following:

E (0–5) K (2–5)
N (0–14) S (2–14).

whereupon the value S can be inserted:

. . . E J G N Z . . .
. . . T . K . . S . . .

k. (1) Consider all the values based upon the displacement interval corresponding to JG:

J (6–5)	G (1–5)→	J (9– 8)	G (4– 8)		
N (6–7)	Z (1–7)	H (9– 9)	O (4– 9)		
		S (9–20)	P (4–20)→	S (2–14)	P (5–14)
				Z (2– 8)	C (5– 8)
				K (2– 5)	D (5– 5)

(2) Since J and G are sequent in the E J G N Z sequence, it can be said that all the letters of the foregoing pairs are also sequent. Hence Z C, S P, and K D are available as new data. These give E J G N Z C and T . K D . S P.

(3) Now consider:

T (0–20)	P (4–20)
A (0– 1)	E (4– 1)
H (0– 8)	G (4– 8)
I (0– 9)	O (4– 9)

Now in the T . K D . S P sequence the interval between T and P is $T\,\overset{1}{.}\,\overset{2}{.}\,\overset{3}{.}\,\overset{4}{.}\,\overset{5}{.}\,\overset{6}{P}$. Hence the interval between A and E is 6 also. It follows therefore that the sequences A V H O F and E J G N Z C should be united, thus:

$$. . . A\;\overset{1}{V}\;\overset{2}{H}\;\overset{3}{O}\;\overset{4}{F}\;\overset{5}{.}\;\overset{6}{E}\;J\;G\;N\;Z\;C\;. . .$$

(4) Corroboration is found in the interval between H and G, which is also six. The letter I can be placed into position, from the relation I (0–9) O (4–9), thus:

$$. . . I\;\overset{1}{.}\;\overset{2}{.}\;\overset{3}{A}\;\overset{4}{V}\;\overset{5}{H}\;\overset{6}{O}\;F\;.\;E\;J\;G\;N\;Z\;C\;. . .$$

l. (1) From Figure 33:

H (0– 8)	Z (2– 8)
E (0– 5)	K (2– 5)
N (0–14)	S (2–14)
U (0–21)	F (2–21)

(2) Since in the I . . A V H O F . E J G N Z C sequence the letters H and Z are separated by 8 intervals one can write:

$$. . . H\;\overset{1}{.}\;\overset{2}{.}\;\overset{3}{.}\;\overset{4}{.}\;\overset{5}{.}\;\overset{6}{.}\;\overset{7}{.}\;\overset{8}{Z}\;. . .$$

. . . E . . . . K . . .
. . . N . . . . S . . .
. . . U . . . . F . . .

(3) Hence one can make the sequence

```
                               1 2 3 4 5 6 7 8
     . . . I . . A V H O F . E J G N Z C . . K . . .
Then . . . I . . A V H O F . E J G N Z C T . K D . S P . . .
and  . . U I . . A V H O F . E J G N Z C T . K D . S P . . .
           1 2 3 4 5 6 7 8           1 2 3 4 5 6 7 8
```

m. (1) Subsequent derivations can be indicated very briefly as follows:

E (0–5)   C (0–3)
D (5–5)   R (5–3)

```
                1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
From            U  I  .  .  A  V  H  O  F  .  E  J  G  N  Z  C  T  .  K  D  .  S  P  .  .  .
one can write                         .  .  .  E  .  .  .  .  C  .  .  .
                                                  1  2  3  4  5
and                                                              .  .  .  D  .  .  .  .  R  .
                                                                              1  2  3  4  5
```

making the sequence

```
                1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
                U  I  .  .  A  V  H  O  F  .  E  J  G  N  Z  C  T  .  K  D  .  S  P  .  R  .
```

(2) Another derivation:

U (3–20)   T (0–20)
X (3– 5)   E (0– 5)

```
                1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
From            U  I  .  .  A  V  H  O  F  .  E  J  G  N  Z  C  T  .  K  D  .  S  P  .  R  .
one can write
                U  I  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T  .  .  .
and             .  .  .  .  .  .  .  .  .  .  E  .  .  .  .  .  .  .  .  .  X  .  .  .  .  .
```

making the sequence

```
                1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
                U  I  .  .  A  V  H  O  F  .  E  J  G  N  Z  C  T  .  K  D  X  S  P  .  R  .
```

(3) Another derivation:

E (0–5)   G (1–5)
B (0–2)   W (1–2)

```
From             . . . E J G . . .
one can write    . . . E . G . . .
and then         . . . B . W . . .
```

There is only one place where B . W can fit, *viz*, at the end:

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
 U  I  .  .  A  V  H  O  F  .  E  J  G  N  Z  C  T  .  K  D  X  S  P  B  R  W
```

n. Only four letters remain to be placed into the sequence, *viz*, L, M, Q, and Y. Their positions are easily found by application of the primary component to the message. The complete sequence is as follows:

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
 U  I  M  Y  A  V  H  O  F  L  E  J  G  N  Z  C  T  Q  K  D  X  S  P  B  R  W
```

Having the primary component fully constructed, decipherment of the cryptogram can be completed with speed and precision. The text is as follows:

```
W F U P C F O C J Y     R C V O P N B L C W     B K D Z F M T G Q J
B U T T H O U G H W     P O S I N G T H E S     S E L F W I L L G O

G B Z D P F B O U O     L Q Z A A A M D C H     L F U Y D T Z V H Q
E C A N N O T A S Y     O L A R S Y S T E M     O U T B E C O M I N

G R F I Z M Q M A V     B Z Z C K Q O I K F     Z G W N K X J T R N
E T R E V I E W W I     S H A L L T U R N A     G A C O L D A N D L

K Z U G D Y F T R W     C F B S C V X C H Q     Y T X C D P M V L W
T H T H E M I N D S     N U N C H A N G I N     I F E L E S S M A S

G J X N L W Y O U X     Z T Z S D M X W C M     B G B W W O Q R G N
E Y E O U R P A S T     G F A C E I N P E R     S A N D T H E S O L

I T W E P Q Z O K Z     R K U H E Q E D G X     H H V L A Q Q V A V
W E C A N T O A N E     P E T U I T Y T O T     A R S Y S T E M W I

P R X C W L Z I C W     F K V H P J J K J Y     J Q W O O T T N V Q
X T E N T F O R E S     H E S U N E A C H W     L L C I R C L E U N

G K Q H O L O D V M     Y Q D P C J X L L L     B K X D S O Z R S N
E E O U R F U T U R     I L L T H E N H A V     S E E N G H O S T L

G O X S N Z H A S E     G H X E R O Q P S E     Y U X O P P Y O X Z
E W E C A N W I T H     E R E A C H E D T H     I K E I N S P A C E

B B J I P Q F J H D     G K B W T L F D U Z     H O Z O W M X C G Q
S C I E N T I F I C     E E N D O F I T S E     A W A I T I N G O N

Q C B Z E X Q T X Z     O C D H W M Z T U Z     J J U G J W Q R V M
C O N F I D E N C E     V O L U T I O N S E     L Y T H E R E S U R

J C Q R Q F V M L M     K L B P C J O T X E     U K W P E F X E N F
L O O K F O R W A R     T I N T H E U N C H     R E C T I O N O F A

S R Q E W M L N A E     H S P O P N M D L M     C C U G D W P E U H
D T O A T I M E W H     A N G I N G S T A R     N O T H E R C O S M

G S X E R O Z J S E     G C K W D V B L S E     Y B W E W V M D Y J
E N E A C H O F T H     E O F D E A T H T H     I C C A T A S T R O

G V Q W E J M K G H     G S U G D P O T H X     R Z X
E B O D I E S C O M     E N T H E S U N I T     P H E
```

FIGURE 34.

o. The primary component appears to be a random-mixed sequence; no key word is to be found, at least none reappears on experimentation with various hypotheses as to enciphering equations. Nevertheless, the random construction of the primary component did not complicate or retard the solution.

p. Some students may prefer to work exclusively with the reconstruction skeleton, rather than with sliding strips. One method is as good as the other and personal preferences will dictate which will be used by the individual student. If the reconstruction skeleton is used, the original letters should be inserted in ink, so as to differentiate them from derived letters.

36. General remarks.—*a*. It is to be stated that the sequence of steps described in the preceding paragraphs corresponds quite closely with that actually followed in solving the problem. It is also to be pointed out that this method can be used as a control in the early stages of analysis because it will allow the cryptanalyst to check assumptions for values. For example, the very first value derived in applying the principles of indirect symmetry to the problem herein described was $H_c=A_p$ in alphabet 1. As a matter of fact the writer had been inclined toward this value, from a study of the frequency and combinations which H_c showed; when the indirect-symmetry method actually substantiated his tentative hypothesis he immediately proceeded to substitute the value given. If he had assigned a different value to H_c, or if he had assumed a letter other than H_c for A_p in that alphabet, the conclusion would immediately follow that either the assumed value for H_c was erroneous, or that one of the values which led to the derivation of $H_c=A_p$ by indirect symmetry was wrong. Thus, these principles aid not only in the systematic and nearly automatic derivation of new values (with only occasional, or incidental references to the actual frequencies of letters), but they also assist very materially in serving as corroborative checks upon the validity of the assumptions already made.

b. Furthermore, while the writer has set forth, in the reconstruction skeleton in Figure 33, a set of 30 values apparently obtained before he began to reconstruct the primary component, this was done for purposes of clarity and brevity in exposition of the principles herein described. As a matter of fact, what he did was to watch very carefully, when inserting values in the reconstruction skeleton to find the very first chance to employ the principles of indirect symmetry; and just as soon as a value could be derived, he substituted the value in the cryptographic text. This is good procedure for two reasons. Not only will it disclose impossible combinations but also it gives opportunity for making further assumptions for values by the addition of the derived values to those previously assumed. Thus, the processes of reconstructing the primary component and finding additional data for the reconstruction proceed simultaneously in an ever-widening circle.

c. It is worth noting that the careful analysis of only 30 cipher equivalents in the reconstruction skeleton shown in Figure 33 results in the derivation of the entire table of secondary alphabets, 676 values in all. And while the elucidation of the method seems long and tedious, in its actual application the results are speedy, accurate, and gratifying in their corroborative effect upon the mental activity of the cryptanalyst.

d. (1) The problem here used as an illustrative case is by no means one that most favorably presents the application and the value of the method, for it has been applied in other cases with much speedier success. For example, suppose that in a cryptogram of 6 alphabets the equivalents of only THE in all 6 alphabets are fairly certain. As in the previous case, it is supposed that the secondary alphabets are obtained by sliding a mixed alphabet against itself. Suppose the secondary alphabets to be as follows:

0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1					B			Q												E						
2					C			L												X						
3					I			V												C						
4					N			P												B						
5					X			O												P						
6					T			Z												V						

FIGURE 35.

(2) Consider the following chain of derivatives arranged diagrammatically:

```
   H (0- 8)  O (5- 8)
   T (0-20)  P (5-20)
   E (0- 5)  X (5- 5)→E (1-20)  X (2-20)
                      Q (1- 8)  L (2- 8)
                      B (1- 5)  C (2- 5)→B (4-20)  C (3-20)
                                         N (4- 5)  I (3- 5)
                                         P (4- 8)  V (3- 8)→

 →P (5-20)  V (6-20)
   O (5- 8)  Z (6- 8)
   X (5- 5)  T (6- 5)→X (2-20)  T (0-20)
                      L (2- 8)  H (0- 8)
                      C (2- 5)  E (0- 5)→C (3-20)  E (1-20)
                                         V (3- 8)  Q (1- 8)
                                         I (3- 5)  B (1- 5)
```

FIGURE 36.

(3) These pairs manifestly all belong to the same displacement interval, and therefore unions can be made immediately. The complete list is as follows:

EX, QL, NI, LH, HO, BC, OZ, CE, TP, PV, XT, VQ, IB

(4) Joining pairs by their common letters, the following sequence is obtained:

. . . N I B C E X T P V Q L H O Z . . .

e. With this as a nucleus the cryptogram can be solved speedily and accurately. When it is realized that the cryptanalyst can assume THE's rather readily in some cases, the value of this principle becomes apparent. When it is further realized that if a cryptogram has sufficient text to enable the THE's to be found easily, it is usually also not at all difficult to make correct assumptions of values for two or three other high-frequency letters, it is clear that the principles of indirect symmetry of position may often be used with gratifyingly quick success to reconstruct the complete primary component.

f. When the probable-word method is combined with the principles of indirect symmetry the solution of a difficult case is often accomplished with astonishing ease and rapidity.

SECTION IX

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, III

Paragraph
Solution of messages enciphered by known primary components........ 37
Solution of repeating-key ciphers in which the identical mixed components proceed in opposite directions........ 38
Solution of repeating-key ciphers in which the primary components are different mixed sequences........ 39
Solution of subsequent messages after the primary components have been recovered........ 40

37. Solution of subsequent messages enciphered by the same primary components.—*a.* In the discussion of the methods of solving repeating-key ciphers using secondary alphabets derived from the sliding of a mixed component against the normal component (Section V), it was shown how subsequent messages enciphered by the same pair of primary components but with different keys could be solved by application of principles involving the completion of the plain-component sequence (paragraphs 23, 24). The present paragraph deals with the application of these same principles to the case where the primary components are identical mixed sequences.

b. Suppose that the following primary component has been reconstructed from the analysis of a lengthy cryptogram:

```
Q U E S T I O N A B L Y C D F G H J K M P R V W X Z
```

A new message exchanged between the same correspondents is intercepted and is suspected of having been enciphered by the same primary components but with a different key. The message is as follows:

```
N F W W P   N O M K I   W P I D S   C A A E T   Q V Z S E

Y O J S C   A A A F G   R V N H D   W D S C A   E G N F P

F O E M T   H X L J W   P N O M K   I Q D B J   I V N H L

T F N C S   B G C R P
```

c. Factoring discloses that the period is 7 letters. The text is transcribed accordingly, and is as follows:

```
N F W W P N O
M K I W P I D
S C A A E T Q
V Z S E Y O J
S C A A A F G
R V N H D W D
S C A E G N F
P F O E M T H
X L J W P N O
M K I Q D B J
I V N H L T F
N C S B G C R
P
```

FIGURE 37.

d. The letters belonging to the same alphabet are then employed as the initial letters of completion sequences, in the manner shown in paragraph 23*e*, using the already reconstructed primary component. The completion diagrams for the first five letters of the first three alphabets are as follows:

ALPHABET 1

N	M	S	V	S
A	P	T	W	T
B	R	I	X	I
L	V	O	Z	O
Y	W	N	Q	N
C	X	A	U	A
D	Z	B	E	B
F	Q	L	S	L
G	U	Y	T	Y
*H	E	C	I	C
J	S	D	O	D
K	T	F	N	F
M	I	G	A	G
P	O	H	B	H
R	N	J	L	J
V	A	K	Y	K
W	B	M	C	M
X	L	P	D	P
Z	Y	R	F	R
Q	C	V	G	V
U	D	W	H	W
E	F	X	J	X
S	G	Z	K	Z
T	H	Q	M	Q
I	J	U	P	U
O	K	E	R	E

ALPHABET 2

F	K	C	Z	C
G	M	D	Q	D
H	P	F	U	F
J	R	G	E	G
K	V	H	S	H
M	W	J	T	J
P	X	K	I	K
R	Z	M	O	M
V	Q	P	N	P
W	U	R	A	R
X	E	V	B	V
Z	S	W	L	W
Q	T	X	Y	X
U	I	Z	C	Z
E	O	Q	D	Q
S	N	U	F	U
T	A	E	G	E
I	B	S	H	S
O	L	T	J	T
N	Y	I	K	I
*A	C	O	M	O
B	D	N	P	N
L	F	A	R	A
Y	G	B	V	B
C	H	L	W	L
D	J	Y	X	Y

ALPHABET 3

W	I	A	S	A
X	O	B	T	B
Z	N	L	I	L
Q	A	Y	O	Y
U	B	C	N	C
E	L	D	A	D
S	Y	F	B	F
T	C	G	L	G
I	D	H	Y	H
O	F	J	C	J
N	G	K	D	K
A	H	M	F	M
B	J	P	G	P
L	K	R	H	R
Y	M	V	S	V
C	P	W	K	W
D	R	X	M	X
F	V	Z	P	Z
G	W	Q	R	Q
H	X	U	V	U
J	Z	E	W	E
K	Q	S	X	X
M	U	T	Z	T
P	E	I	Q	I
R	S	O	U	O
*V	T	N	E	N

FIGURE 38.

e. Examining the successive generatracies to select the ones showing the best assortment of high-frequency letters, those marked in Figure 38 by asterisks are chosen. These are then assembled in columnar fashion and yield the following plain text:

1	2	3	4	5	6	7
H	A	V				
E	C	T				
C	O	N				
I	M	E				
C	O	N				

FIGURE 39.

f. The corresponding key-letters are sought, using enciphering equations $\theta_{k/e}=\theta_{i/p}$; $\theta_{p/p}=\theta_{c/e}$, and are found to be JOU, which suggests the keyword JOURNEY. Testing the key-letters RNEY for alphabets 4, 5, 6, and 7, the following results are obtained:

```
1 2 3 4 5 6 7
J O U R N E Y
-------------
N F W W P N O
H A V E D I R

M K I W P I D
E C T E D S E
```

FIGURE 40.

The message may now be completed with ease. It is as follows:

```
J O U R N E Y       J O U R N E Y
-------------       -------------
H A V E D I R       S A I N C E I
N F W W P N O       P F O E M T H

E C T E D S E       N T H E D I R
M K I W P I D       X L J W P N O

C O N D R E G       E C T I O N O
S C A A E T Q       M K I Q D B J

I M E N T T O       F H O R S E S
V Z S E Y O J       I V N H L T F

C O N D U C T       H O E F A L L
S C A A A F G       N C S B G C R

T H O R O R E       S
R V N H D W D       P

C O N N A I S
S C A E G N F
```

FIGURE 41.

38. Solution of repeating-key ciphers in which the identical mixed components proceed in opposite directions.—The secondary alphabets in this case (paragraph 6, Case B (3) (a) (II) are reciprocal. The steps in solution are essentially the same as in the preceding case (paragraph 28); the principles of indirect symmetry of position can also be applied with the necessary modifications introduced by virtue of the reciprocity existing within the respective secondary alphabets (paragraph 31*p*).

39. Solution of repeating-key ciphers in which the primary components are different mixed sequences.—This is Case B (3) (b) of paragraph 6. The steps in solution are essentially the same as in paragraphs 28 and 31, except that in applying the principles of indirect symmetry of position it is necessary to take cognizance of the fact that the primary components are different mixed sequences (paragraph 31*q*).

40. Solution of subsequent messages after the primary components have been recovered.—*a.* In the case in which the primary components are identical mixed sequences proceeding in opposite directions, as well as in that in which the primary components are different mixed sequences, the solution of subsequent messages[1] is a relatively easy matter. In both cases, however, the student must remember that before the method illustrated in paragraph 37 can be applied it is necessary to convert the cipher letters into their plain-component equivalents before completing the plain-component sequence. From there on, the process of selecting and assembling the proper generatrices is the same as usual.

b. Perhaps an example may be advisable. Suppose the enemy has been found to be using primary components based upon the keyword QUESTIONABLY, the plain component running from left to right, the cipher component in the reverse direction. The following new message has arrived from the intercept station:

```
MVXOX  BZIYZ  NLWZH  OXIEO  OOEPZ
FXSRX  EJBSH  BONAU  RAPZI  NRAMV
XOXAI  JYXWF  KNDOW  JERCU  RALVB
ZAQUW  JWXYI  DGRKD  QBDRM  QECYV
QW
```

1	2	3	4	5	6
M	V	X	O	X	B
Z	I	Y	Z	N	L
W	Z	H	O	X	I
E	O	O	O	E	P
Z	F	X	S	R	X
E	J	B	S	H	B
O	N	A	U	R	A
P	Z	I	N	R	A
M	V	X	O	X	A
I	J	Y	X	W	F
K	N	D	O	W	J
E	R	C	U	R	A
L	V	B	Z	A	Q
U	W	J	W	X	Y
I	D	G	R	K	D
Q	B	D	R	M	Q
E	C	Y	V	Q	W

FIGURE 42.

c. Factoring discloses that the period is 6 and the message is accordingly transcribed into 6 columns, Fig. 42. The letters of these columns are then converted into their plain component equivalents by juxtaposing the two primary components at any point of coincidence, for example $Q_p=Z_c$. The converted letters are shown in Fig. 43. The letters of the individual columns are then used as the initial letters of completion sequences, using the QUESTIONABLY primary sequence. The final step is the selection and assembling of the selected generatrices. The results for the first ten letters of the first three columns are shown below:

1	2	3	4	5	6
O	S	U	M	U	H
Q	P	F	Q	K	G
E	Q	B	M	U	P
W	M	M	M	W	I
Q	Y	U	V	T	U
W	A	H	V	B	H
M	K	J	X	T	J
I	Q	P	K	T	J
O	S	U	M	U	J
P	A	F	U	E	Y
N	K	C	M	E	A
W	T	D	X	T	J
G	S	H	Q	J	Z
X	E	A	E	U	F
P	C	L	T	N	C
Z	H	C	T	O	Z
W	D	F	S	Z	E

FIGURE 43.

---

[1] That is, messages intercepted after the primary components have been reconstructed and enciphered by keys different from those used in the messages upon which the reconstruction of the primary components was accomplished.

COLUMN 1

O	Q	E	W	Q	W	M	I	O	P
N	U	S	X	U	X	P	O	N	R
A	E	T	Z	E	Z	R	N	A	V
B	S	I	Q	S	Q	V	A	B	W
L	T	O	U	T	U	W	B	L	X
Y	I	N	E	I	E	X	L	Y	Z
C	O	A	S	O	S	Z	Y	C	Q
D	N	B	T	N	T	Q	C	D	U
*F	A	L	I	A	I	U	D	F	E
G	B	Y	O	B	O	E	F	G	S
H	L	C	N	L	N	S	G	H	T
J	Y	D	A	Y	A	T	H	J	I
K	C	F	B	C	B	I	J	K	O
M	D	G	L	D	L	O	K	M	N
P	F	H	Y	F	Y	N	M	P	A
R	G	J	C	G	C	A	P	R	B
V	H	K	D	H	D	B	R	V	L
W	J	M	F	J	F	L	V	W	Y
X	K	P	G	K	G	Y	W	X	C
Z	M	R	H	M	H	C	X	Z	D
Q	P	V	J	P	J	D	Z	Q	F
U	R	W	K	R	K	F	Q	U	G
E	V	X	M	V	M	G	U	E	H
S	W	Z	P	W	P	H	E	S	J
T	X	Q	R	S	R	J	S	T	K
I	Z	U	V	Z	V	K	T	I	M

COLUMN 2

S	P	Q	M	Y	A	K	Q	S	A
T	R	U	P	C	B	M	U	T	B
*I	V	E	R	D	L	P	E	I	L
O	W	S	V	F	Y	R	S	O	Y
N	X	T	W	G	C	V	T	N	C
A	Z	I	X	H	D	W	I	A	D
B	Q	O	Z	J	F	X	O	B	F
L	U	N	Q	K	G	Z	N	L	G
Y	E	A	U	M	H	Q	A	Y	H
C	S	B	E	P	J	U	B	C	J
D	T	L	S	R	K	E	L	D	K
F	I	Y	T	V	M	S	Y	F	M
G	O	C	I	W	P	T	C	G	P
H	N	D	O	X	R	I	D	H	R
J	A	F	N	Z	V	O	F	J	V
K	B	G	A	Q	W	N	G	K	W
M	L	H	B	U	X	A	H	M	X
P	Y	J	L	E	Z	B	J	P	Z
R	C	K	Y	S	Q	L	K	R	Q
V	D	M	C	T	U	Y	M	V	U
W	F	P	D	I	E	C	P	W	E
X	G	R	F	O	S	D	R	X	S
Z	H	V	G	N	T	F	V	Z	T
Q	J	W	H	A	I	G	W	O	I
U	K	X	J	B	O	H	X	U	O
E	M	Z	K	L	N	J	Z	E	N

COLUMN 3

U	F	B	M	U	H	J	P	U	F
E	G	L	P	E	J	K	R	E	G
S	H	Y	R	S	K	M	V	S	H
T	J	C	V	T	M	P	W	T	J
I	K	D	W	I	P	R	X	I	K
O	M	F	X	O	R	V	Z	O	M
N	P	G	Z	N	V	W	Q	N	P
A	R	H	Q	A	W	X	U	A	R
B	V	J	U	B	X	Z	E	B	V
L	W	K	E	L	Z	Q	S	L	W
Y	X	M	S	Y	Q	U	T	Y	X
C	Z	P	T	C	U	E	I	C	Z
D	Q	R	I	D	E	S	O	D	Q
F	U	V	O	F	S	T	N	F	U
G	E	W	N	G	T	I	A	G	E
H	S	X	A	H	I	O	B	H	S
J	T	Z	B	J	O	N	L	J	T
K	I	Q	L	K	N	A	Y	K	I
M	O	U	Y	M	A	B	C	M	O
P	N	E	C	P	B	L	D	P	N
*R	A	S	D	R	L	Y	F	R	A
V	B	T	F	V	Y	C	G	V	B
W	L	I	G	W	C	D	H	W	L
X	Y	O	H	X	D	F	J	X	Y
Z	C	N	J	Z	F	G	K	Z	C
Q	D	A	K	Q	G	H	N	Q	D

FIGURE 44.

Columnar assembling of selected generatrices gives what is shown in Fig. 45.

1	2	3	4	5	6
F	I	R	.	.	.
A	V	A	.	.	.
L	E	S	.	.	.
I	R	D	.	.	.
A	D	R	.	.	.
I	L	L	.	.	.
U	P	Y	.	.	.
D	E	F	.	.	.
F	I	R	.	.	.
E	L	A	.	.	.

FIGURE 45.

d. The key letters are sought, and found to be NUM, which suggests NUMBER. The entire message may now be read with ease. It is as follows:

N	U	M	B	E	R
F	I	R	S	T	C
M	V	X	O	X	B
A	V	A	L	R	Y
Z	I	Y	Z	N	L
L	E	S	S	T	H
W	Z	H	O	X	I
I	R	D	S	Q	U
E	O	O	O	E	P
A	D	R	O	N	W
Z	F	X	S	R	X
I	L	L	O	C	C
E	J	B	S	H	B
U	P	Y	A	N	D
O	N	A	U	R	A
D	E	F	E	N	D
P	Z	I	N	R	A
F	I	R	S	T	D
M	V	X	O	X	A

N	U	M	B	E	R
E	L	A	Y	I	N
I	J	Y	X	W	F
G	P	O	S	I	T
K	N	D	O	W	J
I	O	N	A	N	D
E	R	C	U	R	A
W	I	L	L	P	R
L	V	B	Z	A	Q
O	T	E	C	T	L
U	W	J	W	X	Y
E	F	T	F	L	A
I	D	G	R	K	D
N	K	O	F	B	R
Q	B	D	R	M	Q
I	G	A	D	E	X
E	C	Y	V	Q	W

FIGURE 46.

e. If the primary components are different mixed sequences, the procedure is identical with that just indicated. The important point to note is that one must not fail to convert the letters into their plain-component equivalents before the completion-sequence method is applied.

SECTION X

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, IV

Paragraph
General remarks 41
Deriving the secondary alphabets, the primary components, and the key, given a cryptogram with its plain text 42
Deriving the secondary alphabets, the primary components, and the keywords for messages, given two or more cryptograms in different keys and suspected to contain identical plain text 43
The case of repeating-key systems 44
The case of identical messages enciphered by keywords of different lengths 45
Concluding remarks 46

41. General remarks.—The preceding three sections have been devoted to an elucidation of the general principles and procedure in the solution of typical cases of repeating-key ciphers. This section will be devoted to a consideration of the variations in cryptanalytic procedure arising from special circumstances. It may be well to add that by the designation "special circumstances" it is not meant to imply that the latter are necessarily *unusual* circumstances. *The student should always be on the alert to seize upon any opportunities that may appear in which he may apply the methods to be described.* In practical work such opportunities are by no means rare and are seldom overlooked by competent cryptanalysts.

42. Deriving the secondary alphabets, the primary components, and the key, given a cryptogram with its plain text.—*a.* It may happen that a cryptogram and its equivalent plain text are at hand, as the result of capture, pilferage, compromise, etc. This, as a general rule, affords a very easy attack upon the whole system.

b. Taking first the case where the plain component is the normal alphabet, the cipher component a mixed sequence, the first thing to do is to write out the cipher text with its letter-for-letter decipherment. From this, by a slight modification of the principles of "factoring", one discovers the length of the key. It is obvious that when a word of three or four letters is enciphered by the same cipher text, the interval between the two occurrences is almost certainly a multiple of the length of the key. By noting a few recurrences of plain text and cipher letters, one can quickly determine the length of the key (assuming of course that the message is long enough to afford sufficient data). Having determined the length of the key, the message is rewritten according to its periods, with the plain text likewise in periods under the cipher letters. From this arrangement one can now reconstruct complete or partial secondary alphabets. If the secondary alphabets are complete, they will show direct symmetry of position; if they are but fragmentary in several alphabets, then the primary component can be reconstructed by the application of the principles of direct symmetry of position.

c. If the plain component is a mixed sequence, and the cipher component the normal (direct or reversed sequence), the secondary alphabets will show no direct symmetry unless they are arranged in the form of deciphering alphabets (that is, $A_e \ldots Z_e$ above the zero line, with their equivalents below). The student should be on the lookout for such cases.

d. (1) If the plain and cipher primary components are identical mixed sequences proceeding in the same direction, the secondary alphabets will show indirect symmetry of position, and they can be used for the speedy reconstruction of the primary components (Paragraph 31*a* to *o*).

(2) If the plain and the cipher primary components are identical mixed sequences proceeding in opposite directions, the secondary alphabets will be completely reciprocal secondary alphabets and the primary component may be reconstructed by applying the principles outlined in paragraph 31*p*.

(3) If the plain and the cipher primary components are different mixed sequences, the secondary alphabets will show indirect symmetry of position and the primary components may be reconstructed by applying the principles outlined in paragraph 31*q*.

e. In all the foregoing cases, after the primary components have been reconstructed, the keys can be readily recovered.

43. Deriving the secondary alphabets, the primary components, and the keywords for messages, given two or more cryptograms in different keys and suspected to contain identical plain text.—*a*. The simplest case of this kind is that involving two monoalphabetic substitution ciphers with mixed alphabets derived from the same pair of sliding components. An understanding of this case is necessary to that of the case involving repeating-key ciphers.

b. (1) A message is transmitted from station A to station B. B then sends A some operating signals which indicate that B cannot decipher the message, and soon thereafter A sends a second message, identical in length with the first. This leads to the suspicion that the plain text of both messages is the same. The intercepted messages are superimposed. Thus:

```
1. NXGRV MPUOF ZQVCP VWERX QDZVX WXZQE TBDSP VVXJK RFZWH ZUWLU IYVZQ FXOAR
2. EMLHJ FGVUB PRJNG JKWHM RAPJM KMPRW ZTAXG JJMCD HBPKY PVKIV QOJPR BMUSH
```

(2) Initiating a chain of cipher-text equivalents from message 1 to message 2, the following complete sequence is obtained:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
N	E	W	K	D	A	S	X	M	F	B	T	Z	P	G	L	I	Q	R	H	Y	O	U	V	J	C

(3) Experimentation along already-indicated lines soon discloses the fact that the foregoing component is an equivalent primary component of the original primary based upon the keyword QUESTIONABLY, decimated on the 21st interval. Let the student decipher the cryptogram.

(4) The foregoing example is somewhat artificial in that the plain text was consciously selected with a view to making it contain every letter of the alphabet. The purpose in doing this was to permit the construction of a complete chain of equivalents from only two short messages, in order to give a simple illustration of the principles involved. If the plain-text message does not contain every letter of the alphabet, then only partial chains of equivalents can be constructed. These may be united, if circumstances will permit, by recourse to the various principles elucidated in paragraph 31.

(5) The student should carefully study the foregoing example in order to obtain a thorough comprehension of the *reason* why it was possible to reconstruct the primary component from the two cipher messages without having any plain text to begin with at all. Since the plain text of both messages is the same, the relative displacement of the primary components in the case of message 1 differs from the relative displacement of the same primary components in the case of message 2 by a *fixed* interval. Therefore, the distance between N and E (the first letters of the two messages), on the primary component, regardless of what plain-text letter these two cipher letters represent, is the same as the distance between E and W (the 18th letters), W and K (the 17th letters), and so on. Thus, this fixed interval permits of establishing a complete chain of letters separated by constant intervals and this chain becomes an equivalent primary component.

44. The case of repeating-key systems.—*a.* With the foregoing basic principles in mind the student is ready to note the procedure in the case of two repeating-key ciphers having identical plain texts. First, the case in which both messages have keywords of identical length but different compositions will be studied.

b. (1) Given the following two cryptograms suspected to contain the same plain text:

Message 1

Y H Y E X U B U K A P V L L T A B U V V D Y S A B
P C Q T U N G K F A Z E F I Z B D J E Z A L V I D
T R O Q S U H A F K

Message 2

C G S L Z Q U B M N C T Y B V H L Q F T F L R H L
M T A I Q Z W M D Q N S D W N L C B L Q N E T O C
V S N Z R B J N O Q

(2) The first step is to try to determine the length of the period. The usual method of factoring cannot be employed because there are no long repetitions and not enough repetitions even of digraphs to give any convincing indications. However, a subterfuge will be employed, based upon the theory of factoring.

c. (1) Let the two messages be superimposed.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1.	Y	H	Y	E	X	U	B	U	K	A	P	V	L	L	T	A	B	U	V	V
2.	C	G	S	L	Z	Q	U	B	M	N	C	T	Y	B	V	H	L	Q	F	T

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1.	D	Y	S	A	B	P	C	Q	T	U	N	G	K	F	A	Z	E	F	I	Z
2.	F	L	R	H	L	M	T	A	I	Q	Z	W	M	D	Q	N	S	D	W	N

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1.	B	D	J	E	Z	A	L	V	I	D	T	R	O	Q	S	U	H	A	F	K
2.	L	C	B	L	Q	N	E	T	O	C	V	S	N	Z	R	B	J	N	O	Q

(2) Now let a search be made of cases of identical superimposition. For example, $\overset{\overset{4}{E}}{L}$ and $\overset{\overset{44}{E}}{L}$ are separated by 40 letters, $\overset{\overset{6}{U}}{Q}$, $\overset{\overset{18}{U}}{Q}$, and $\overset{\overset{30}{U}}{Q}$ are separated by 12 letters. Let these intervals between identical superimpositions be factored, just as though they were ordinary repetitions. That factor which is the most frequent should correspond with the length of the period for the following reason. If the period is the same and the plain text is the same in both messages, then the condition of identity of superimposition can only be the result of identity of encipherments by identical cipher alphabets. This is only another way of saying that the same relative position in the keying cycle has been reached in both cases of identity. Therefore, the distance between identical superimpositions must be either equal to or else a multiple of the length of the period. Hence, factoring the intervals must yield the length of the period. The complete list of intervals

and factors applicable to cases of identical superimposed pairs is as follows (factors above 12 are omitted):

Repetition	Interval	Factors	Repetition	Interval	Factors
1st EL to 2d EL	40	2, 4, 5, 8, 10.	1st TV to 2d TV	36	2, 3, 4, 6, 9, 12.
1st UQ to 2d UQ	12	2, 3, 4, 6, 12.	1st AH to 2d AH	8	2, 4, 8.
2d UQ to 3d UQ	12	2, 3, 4, 6, 12.	1st BL to 2d BL	8	2, 4, 8.
1st UB to 2d UB	48	2, 3, 4, 6, 8, 12.	2d BL to 3d BL	16	2, 4, 8.
1st KM to 2d KM	24	2, 3, 4, 6, 8, 12.	1st SR to 2d SR	32	2, 4, 8.
1st AN to 2d AN	36	2, 3, 4, 6, 9, 12.	1st FD to 2d FD	4	2, 4.
2d AN to 3d AN	12	2, 3, 4, 6, 12.	1st ZN to 2d ZN	4	2, 4.
1st VT to 2d VT	8	2, 4, 8.	1st DC to 2d DC	8	2, 4, 8.
2d VT to 3d VT	28	2, 4, 7.			

(3) The factors 4 and 2 are the only ones common to every one of these intervals and since a period of 2 is not very probable it may be taken as beyond question that the length of the period is 4.

d. Let the messages now be superimposed according to their periods:

```
   1 2 3 4   1 2 3 4   1 2 3 4   1 2 3 4   1 2 3 4   1 2 3 4   1 2 3 4
1. Y H Y E   X U B U   K A P V   L L T A   B U V V   D Y S A   B P C Q
2. C G S L   Z Q U B   M N C T   Y B V H   L Q F T   F L R H   L M T A

1. T U N G   K F A Z   E F I Z   B D J E   Z A L V   I D T R   O Q S U
2. I Q Z W   M D Q N   S D W N   L C B L   Q N E T   O C V S   N Z R B

1. H A F K
2. J N O Q
```

e. (1) Now distribute the superimposed letters into a reconstruction skeleton of "secondary alphabets."
Thus:

0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1		L		F	S			J	O		M	Y			N					I				Z	C	Q
2	N			C		D		G				B				M	Z				Q				L	
3	Q	U	T			O			W	B		E		Z		C			R	V		F			S	
4	H				L		W				Q						A	S			B	T				N

(2) By the usual methods, construct the primary or an equivalent primary component. Taking lines 0 and 1, the following sequences are noted:

BL, DF, ES, HJ, IO, KM, LY, ON, TI, XZ, YC, ZQ,

which, when united by means of common letters and study of other sequences, yield the complete original primary component based upon the keyword QUESTIONABLY:

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z

(3) The fact that the pair of lines with which the process was commenced yield the original primary sequence is purely accidental; it might have just as well yielded an equivalent primary sequence.

f. (1) Having the primary component, the solution of the messages is now a relatively simple matter. An application of the method elucidated in paragraph 37 is made, involving the completion of the plain-component sequence for each alphabet and selecting those generatrices which contain the best assortments of high-frequency letters. Thus, using Message 1:

FIRST ALPHABET	SECOND ALPHABET	THIRD ALPHABET	FOURTH ALPHABET
Y X K L B	H U A L U	Y B P T V	E U V A V
C Z M Y L	J E B Y E	C L R I W	S E W B W
D Q P C Y	K S L C S	D Y V O X	T S X L X
F U R D C	M T Y D T	F C W N Z	I T Z Y Z
G E V F D	P I C F I	G D X A Q	O I Q C Q
H S W G F	R O D G O	H F Z B U	N O U D U
J T X H G	V N F H N	J G Q L E	*A N E F E
K I Z J H	W A G J A	K H U Y S	B A S G S
M O Q K J	X B H K B	M J E C T	L B T H T
P N U M K	Z L J M L	P K S D I	Y L I J I
R A E P M	Q Y K P Y	R M T F O	C Y O K O
V B S R P	U C M R C	V P I G N	D C N M N
W L T V R	E D P V D	W R O H A	F D A P A
X Y I W V	S F R W F	X V N J B	G F B R B
Z C O X W	T G V X G	Z W A K L	H G L V L
Q D N Z X	I H W Z H	Q X B M Y	J H Y W Y
U F A Q Z	O J X Q J	U Z L P C	K J C X C
E G B U Q	N K Z U K	E Q Y R D	M K D Z D
S H L E U	A M Q E M	S U C V F	P M F Q F
T J Y S E	B P U S P	T E D W G	R P G U G
I K C T S	*L R E T R	I S F X H	V R H E H
O M D I T	Y V S I V	O T G Z J	W V J S J
N P F O I	C W T O W	N I H Q K	X W K T K
*A R G N O	D X I N X	A O J U M	Z X M I M
B V H A N	F Z O A Z	B N K E P	Q Z P O P
L W J B A	G Q N B Q	*L A M S R	U Q R N R

FIGURE 48.

(2) The selected generatrices (those marked by asterisks in Fig. 48) are assembled in columnar manner:

A L L A
R R A N
G E M E
N T S F
O R R E

FIGURE 49.

(3) The key letters are sought and give the keyword SOUP. The plain text for the second message is now known, and by reference to the cipher text and the primary components, the keyword for this message is found to be TIME. The complete texts are as follows:

S O U P	T I M E
A L L A	A L L A
Y H Y E	C G S L
R R A N	R R A N
X U B U	Z Q U B
G E M E	G E M E
K A P V	M N C T
N T S F	N T S F
L L T A	Y B V H
O R R E	O R R E
B U V V	L Q F T
L I E F	L I E F
D Y S A	F L R H
O F Y O	O F Y O
B P C Q	L M T A
U R O R	U R O R
T U N G	I Q Z W
G A N I	G A N I
K F A Z	M D Q N
Z A T I	Z A T I
E F I Z	S D W N
O N H A	O N H A
B D J E	L C B L
V E B E	V E B E
Z A L V	Q N E T
E N S U	E N S U
I D T R	O C V S
S P E N	S P E N
O Q S U	N Z R B
D E D X	D E D X
H A F K	J N O Q

FIGURE 50.

45. The case of identical messages enciphered by keywords of different lengths.—*a*. In the foregoing case the keywords for the two messages, although different, were identical in length. When this is not true and the keywords are of different lengths, the procedure need be only slightly modified.

b. Given the following two cryptograms suspected of containing the same plain-text enciphered by the same primary components but with different keywords of different lengths, solve the messages.

MESSAGE No. 1

VMYZG EAUNT PKFAY JIZMB UMYKB VFIVV
SEOAF SKXKR YWCAC ZORDO ZRDEF BLKFE
SMKSF AFEKV QURCM YZVOX VABTA YYUOA
YTDKF ENWNT DBQKU LAJLZ IOUMA BOAFS
KXQPU YMJPW QTDBT OSIYS MIYKU ROGMW
CTMZZ VMVAJ

MESSAGE No. 2

ZGANW IOMOA CODHA CLRLP MOQOJ EMOQU
DHXBY UQMGA UVGLQ DBSPU OABIR PWXYM
OGGFT MRHVF GWKNI VAUPF ABRVI LAQEM
ZDJXY MEDDY BOSVM PNLGX XDYDO PXBYU
QMNKY FLUYY GVPVR DNCZE KJQOR WJXRV
GDKDS XCEEC

c. The messages are long enough to show a few short repetitions which permit factoring. The latter discloses that Message 1 has a period of 4 and Message 2, a period of 6 letters. The messages are superimposed, with numbers marking the position of each letter in the corresponding period, as shown below:

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. V M Y Z G E A U N T P K F A Y J I Z M B U M Y K
No. 2. Z G A N W I O M O A C O D H A C L R L P M O Q O
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. B V F I V V S E O A F S K X K R Y W C A C Z O R
No. 2. J E M O Q U D H X B Y U Q M G A U V G L Q D B S
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. D O Z R D E F B L K F E S M K S F A F E K V Q U
No. 2. P U O A B I R P W X Y M O G G F T M R H V F G W
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. R C M Y Z V O X V A B T A Y Y U O A Y T D K F E
No. 2. K N I V A U P F A B R V I L A Q E M Z D J X Y M
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. N W N T D B Q K U L A J L Z I O U M A B O A F S
No. 2. E D D Y B O S V M P N L G X X D Y D O P X B Y U
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. K X Q P U Y M J P W Q T D B T O S I Y S M I Y K
No. 2. Q M N K Y F L U Y Y G V P V R D N C Z E K J Q O
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. U R O G M W C T M Z Z V M V A J
No. 2. R W J X R V G D K D S X C E E C
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4

d. A reconstruction skeleton of "secondary alphabets" is now made by distributing the letters in respective lines corresponding to the 12 different superimposed pairs of numbers. For example, all pairs corresponding to the superimposition of position 1 of Message 1 with position 1 of Message 2 are distributed in lines Ø and 1 of the skeleton. Thus, the very first superimposed pair is $\left\{\begin{matrix}1\\ \mathrm{V}\\ \mathrm{Z}\\ 1\end{matrix}\right.$; the letter Z is inserted in line 1 under the letter V. The next $\left\{\begin{smallmatrix}1\\ 1\end{smallmatrix}\right.$ pair is the 13th superimposition, with $\left\{\begin{matrix}\mathrm{F}\\ \mathrm{D}\end{matrix}\right.$; the letter D is inserted in line 1 under the letter F, and so on. The skeleton is then as follows:

	Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Superimposed pairs.	1–1	I	J		P		D					Q	G	C	E				K	O		R	Z				
	2–2	H	V	N										G		U			W				E	D	M	L	X
	3–3	E					M			X		G		I	D	J		N			R					A	O
	4–4							X		O	C					D	K		A	F	Y	Q				V	N
	1–5				B		T	W		L				R		E				N		Y	Q			U	A
	2–6	M	O			I				C				D									U	V		F	R
	3–1	O		G			R							L		P		S		D						Z	
	4–2	L	P			H					U	V								E	D	M			F		
	1–3			Q	J							V	W	K	O	X	Y					M	A				
	2–4	B								J		X	P	O							A		F	Y			D
	3–5	N	R				Y									B	C	G								Q	S
	4–6					M					L	O							S	U	V	W	X				

FIGURE 51.

e. There are more than sufficient data here to permit of the reconstruction of a complete equivalent primary component, for example, the following:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
I	T	K	N	P	Z	H	M	W	B	Q	E	U	L	F	C	S	J	A	X	R	G	D	V	O	Y

f. The subsequent steps in the actual decipherment of the text of either of the two messages are of considerable interest. Thus far the cryptanalyst has only the cipher component of the primary sliding components. The plain component may be identical with the cipher component and may progress in the same direction, or in the reverse direction; or, the two components may be different. If different, the plain component may be the normal sequence, direct or reversed. Tests must be made to ascertain which of these various possibilities is true.

g. (1) It will first be assumed that the primary plain component is the normal direct sequence. Applying the procedure outlined in Par. 23 to the message with the shorter key (Message No. 1, to give the most data per secondary alphabet), an attempt is made to solve the message. It is unnecessary here to go further into detail in this procedure; suffice it to indicate that the attempt is unsuccessful and it follows that the plain component is not the normal direct sequence. A normal reversed sequence is then assumed for the plain component and the proper procedure applied. Again the attempt is found useless. Next, it is assumed that the plain component is identical with the cipher component, and the procedure outlined in Par. 37 is tried. This also is unsuccessful. Another attempt, assuming the plain component runs in the reverse direction, is likewise unsuccessful. There remains one last hypothesis, viz, that the two primary components are different mixed sequences.

(2) Here is Message No. 1 transcribed in periods of four letters. Uniliteral frequency distributions for the four secondary alphabets are shown below in Fig. 52, labeled 1*a*, 2*a*, 3*a*, and 4*a*. These distributions are based upon the normal sequence A to Z. But since the reconstructed cipher component is at hand these distributions can be rearranged according to the sequence of the cipher component, as shown in distributions labeled 1*b*, 2*b*, 3*b*, and 4*b* in Fig. 52. *The latter distributions may be combined by shifting distributions 2b, 3b, and 4b to proper superimpositions with respect to 1b so as to yield a single monoalphabetic distribution for the entire message. In other words, the polyalphabetic message can be converted into monoalphabetic terms, thus very considerably simplifying the solution.*

MESSAGE No. 1

VMYZ VABT
GEAU AYYU
NTPK OAYT
FAYJ DKFE
IZMB NWNT
UMYK DBQK
BVFI ULAJ
VVSE LZIO
OAFS UMAB
KXKR OAFS
YWCA KXQP
CZOR UYMJ
DOZR PWQT
DEFB DBTO
LKFE SIYS
SMKS MIYK
FAFE UROG
KVQU MWCT
RCMY MZZV
ZVOX MVAJ

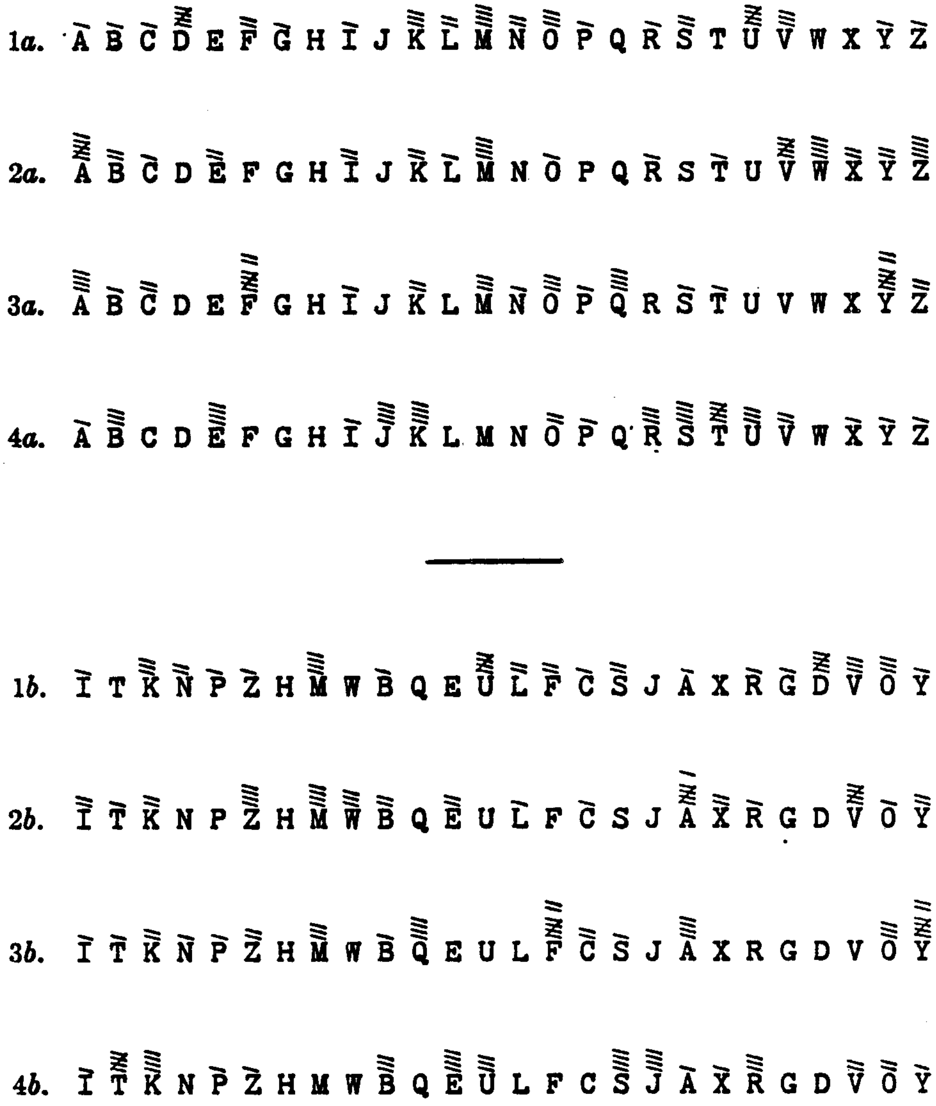


FIGURE 52.

(3) Note in Fig. 53 how the four distributions are shifted for superimposition and how the combined distribution presents the characteristics of a typical monoalphabetic distribution.

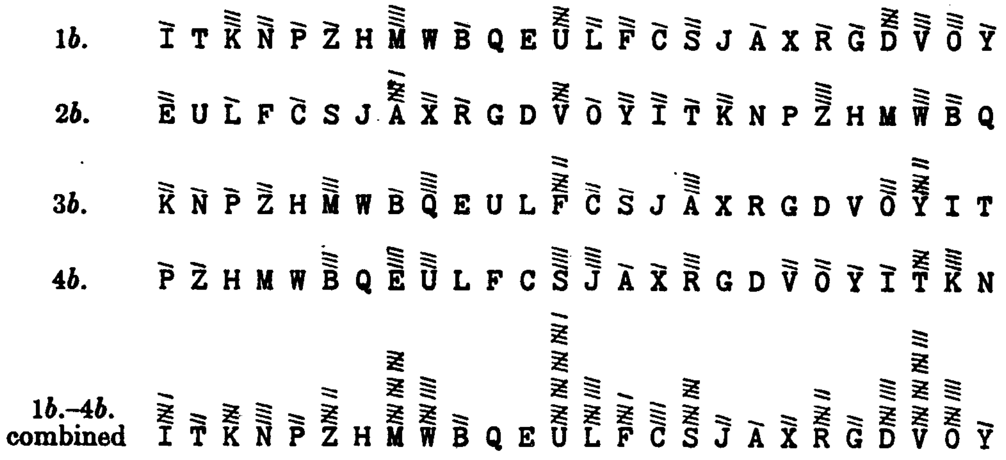


FIGURE 53.

(4) The letters belonging to alphabets 2, 3, and 4 of Fig. 52 may now be transcribed in terms of alphabet 1. That is, the two E's of alphabet 2 become I's; the L of alphabet 2 becomes a K; the C becomes a P, and so on. Likewise, the two K's of alphabet 3 become I's, the N becomes a T, and so on. The entire message is then a monoalphabet and can readily be solved. It is as follows:

V D V T G   I S W N S   K O F M V   L I R Z Z   U D V O B   U U D V U
E N E M Y   H A S C A   P T U R E   D H I L L   O N E T W   O O N E O

F M O M U   U K W I S   Y V L F C   R D S D L   N S D I U   Z L J U M
U R T R O   O P S H A   V E D U G   I N A N D   C A N H O   L D F O R

S D I U F   M U M K U   W W R P Z   G Z U D C   V M M V A   F V W O M
A N H O U   R O R P O   S S I B L   Y L O N G   E R R E Q   U E S T R

V V D J U   M N V T V   D O W O U   K S L L R   O R U D S   Z O M U U
E I N F O   R C E M E   N T S T O   P A D D I   T I O N A   L T R O O

K W W I U   F Z L P V   W V D O Y   R S C V U   M C V O U   B D J M V
P S S H O   U L D B E   S E N T V   I A G E O   R G E T O   W N F R E

L V M R N   X M U S L
D E R I C   K R O A D

(5) Having the plain text, the derivation of the plain component (an equivalent) is an easy matter. It is merely necessary to base the reconstruction upon any of the secondary alphabets, since the plain text—cipher relationship is now known directly, and the primary cipher component is at hand. The primary plain component is found to be as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	M	P	C	B	L	.	R	S	W	.	.	O	D	U	G	A	F	Q	K	I	Y	N	E	T	V

(6) The keywords for both messages can now be found, if desirable, by finding the equivalent of A_p in each of the secondary alphabets of the original polyalphabetic messages. The keyword for No. 1 is STAR; that for No. 2 is OCEANS.

(7) The student may, if he wishes, try to find out whether the primary components reconstructed above are the original components or are equivalent components, by examining all the possible decimations of the two components for evidences of derivation from keywords.

h. As already stated in Par. 26*m*, there are certain statistical and mathematical tests that can be employed in the process of "matching" distributions to ascertain proper superimpositions for monoalphabeticity. In the case just considered there were sufficient data in the distributions to permit the process to be applied successfully by eye, without necessitating statistical tests.

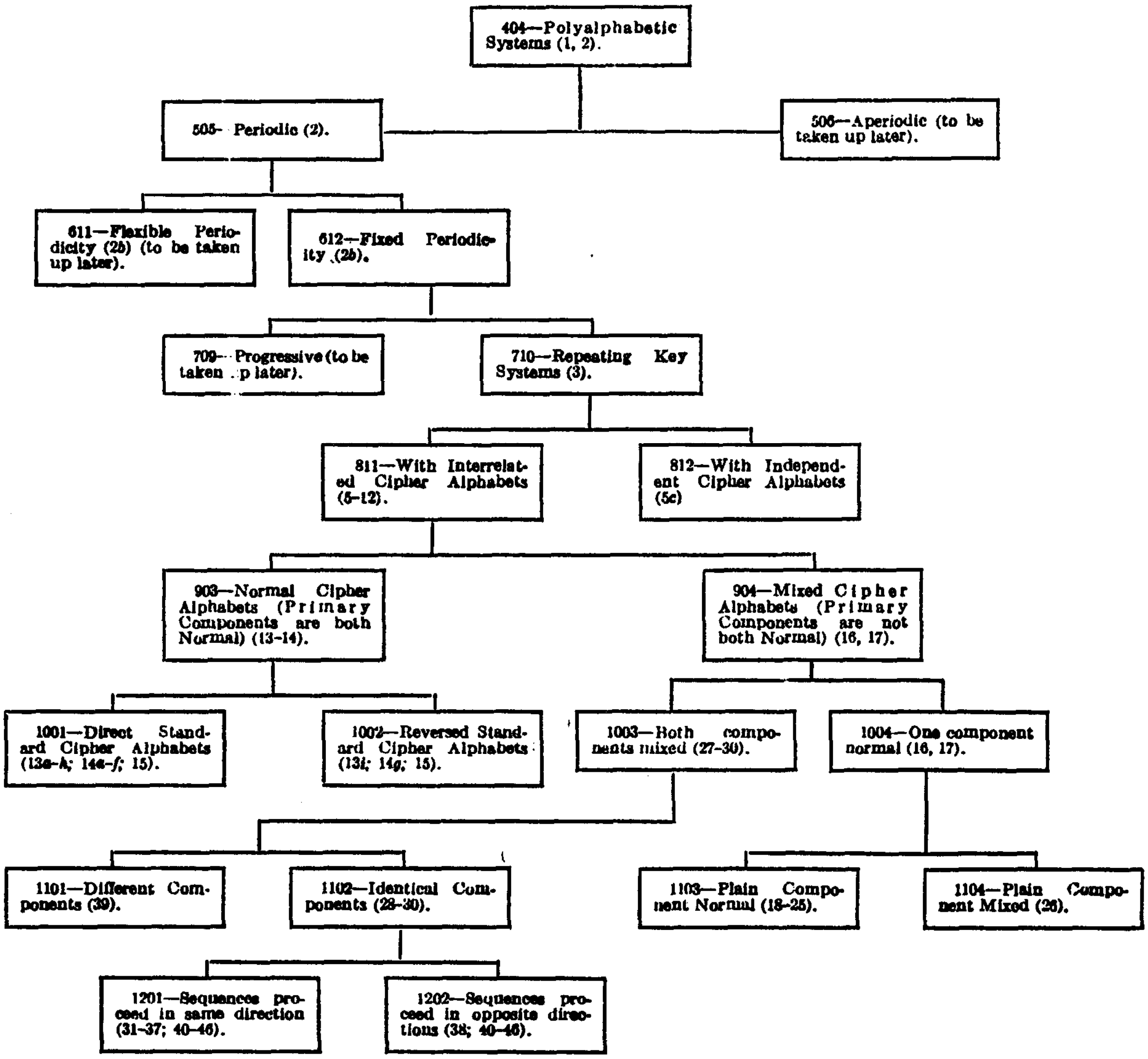
i. This case is an excellent illustration of the application of the process of *converting a polyalphabetic cipher into monoalphabetic terms.* Because it is a very valuable and important cryptanalytic "trick," the student should study it most carefully in order to gain a good understanding of the principle upon which it is based and its significance in cryptanalysis. The conversion in the case under discussion was possible because the sequence of letters forming the cipher component had been reconstructed and was known, and therefore the uniliteral distributions for the respective secondary cipher alphabets could theoretically be shifted to correct superimpositions for monoalphabeticity. It also happened that there were sufficient data in the distributions to give proper indications for their relative displacements. Therefore, the theoretical possibility in this case became an actuality. Without these two necessary conditions the superimposition and conversion cannot be accomplished. The student should always be on the lookout for situations in which this is possible.

46. Concluding remarks.—*a.* The observant student will have noted that a large part of this text is devoted to the elucidation and application of a very few basic principles. These principles are, however, extremely important and their proper usage in the hands of a skilled cryptanalyst makes them practically indispensable tools of his art. The student should therefore drill himself in the application of these tools by having someone make up problem after problem for him to practice upon, until he acquires facility in their use and feels competent to apply them in practice whenever the least opportunity presents itself. This will save him much time and effort in the solution of bona fide messages.

b. Continuing the analytical key introduced in Military Cryptanalysis Part I, the outline for the studies covered by Part II follows herewith.

Analytical Key for Military Cryptanalysis, Part II *

(Numbers in parentheses refer to Paragraph Numbers in this text)


404—Polyalphabetic Systems (1, 2).
505—Periodic (2).
506—Aperiodic (to be taken up later).
611—Flexible Periodicity (2b) (to be taken up later).
612—Fixed Periodicity (2b).
709—Progressive (to be taken up later).
710—Repeating Key Systems (3).
811—With Interrelated Cipher Alphabets (5-12).
812—With Independent Cipher Alphabets (5c)
903—Normal Cipher Alphabets (Primary Components are both Normal) (13-14).
904—Mixed Cipher Alphabets (Primary Components are not both Normal) (16, 17).
1001—Direct Standard Cipher Alphabets (13a-h; 14a-f; 15).
1002—Reversed Standard Cipher Alphabets (13i; 14g; 15).
1003—Both components mixed (27-30).
1004—One component normal (16, 17).
1101—Different Components (39).
1102—Identical Components (28-30).
1103—Plain Component Normal (18-25).
1104—Plain Component Mixed (26).
1201—Sequences proceed in same direction (31-37; 40-46).
1202—Sequences proceed in opposite directions (38; 40-46).

*For explanation of the use of this chart see Par. 50 of Military Cryptanalysis, Part I.

APPENDIX 1

THE 12 TYPES OF CIPHER SQUARES

(See Paragraph 7)

TABLE I–B.[1]

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/2}=\theta_{1/1}$; $\theta_{p/1}=\theta_{c/2}$ ($\theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V
B	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F
C	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R
D	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T
E	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S
F	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X
G	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I
H	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E
I	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z
J	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D
K	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M
L	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A
M	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U
N	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W
O	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N
P	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B
Q	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C
R	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y
S	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G
T	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H
U	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J
V	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K
W	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L
X	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O
Y	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P
Z	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q

---

[1] This table is labeled "Table 1–B" because it is the same as Table 1–A on page 7, except that the horizontal lines of the latter have been shifted so as to begin the successive alphabets with the successive letters of the normal sequence.

TABLE II

Components:

(1)	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
(2)	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X

Enciphering equations: $\theta_{k/2}=\theta_{1/1}$; $\theta_{p/2}=\theta_{c/1}$ ($\theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	H	L	U	R	G	P	S	O	V	Y	B	X	D	E	I	M	K	Q	T	W	Z	C	F	J	N
B	T	A	E	N	K	Z	I	L	H	O	R	U	Q	W	X	B	F	D	J	M	P	S	V	Y	C	G
C	P	W	A	J	G	V	E	H	D	K	N	Q	M	S	T	X	B	Z	F	I	L	O	R	U	Y	C
D	G	N	R	A	X	M	V	Y	U	B	E	H	D	J	K	O	S	Q	W	Z	C	F	I	L	P	T
E	J	Q	U	D	A	P	Y	B	X	E	H	K	G	M	N	R	V	T	Z	C	F	I	L	O	S	W
F	U	B	F	O	L	A	J	M	I	P	S	V	R	X	Y	C	G	E	K	N	Q	T	W	Z	D	H
G	L	S	W	F	C	R	A	D	Z	G	J	M	I	O	P	T	X	V	B	E	H	K	N	Q	U	Y
H	I	P	T	C	Z	O	X	A	W	D	G	J	F	L	M	Q	U	S	Y	B	E	H	K	N	R	V
I	M	T	X	G	D	S	B	E	A	H	K	N	J	P	Q	U	Y	W	C	F	I	L	O	R	V	Z
J	F	M	Q	Z	W	L	U	X	T	A	D	G	C	I	J	N	R	P	V	Y	B	E	H	K	O	S
K	C	J	N	W	T	I	R	U	Q	X	A	D	Z	F	G	K	O	M	S	V	Y	B	E	H	L	P
L	Z	G	K	T	Q	F	O	R	N	U	X	A	W	C	D	H	L	J	P	S	V	Y	B	E	I	M
M	D	K	O	X	U	J	S	V	R	Y	B	E	A	G	H	L	P	N	T	W	Z	C	F	I	M	Q
N	X	E	I	R	O	D	M	P	L	S	V	Y	U	A	B	F	J	H	N	Q	T	W	Z	C	G	K
O	W	D	H	Q	N	C	L	O	K	R	U	X	T	Z	A	E	I	G	M	P	S	V	Y	B	F	J
P	S	Z	D	M	J	Y	H	K	G	N	Q	T	P	V	W	A	E	C	I	L	O	R	U	X	B	F
Q	O	V	Z	I	F	U	D	G	C	J	M	P	L	R	S	W	A	Y	E	H	K	N	Q	T	X	B
R	Q	X	B	K	H	W	F	I	E	L	O	R	N	T	U	Y	C	A	G	J	M	P	S	V	Z	D
S	K	R	V	E	B	Q	Z	C	Y	F	I	L	H	N	O	S	W	U	A	D	G	J	M	P	T	X
T	H	O	S	B	Y	N	W	Z	V	C	F	I	E	K	L	P	T	R	X	A	D	G	J	M	Q	U
U	E	L	P	Y	V	K	T	W	S	Z	C	F	B	H	I	M	Q	O	U	X	A	D	G	J	N	R
V	B	I	M	V	S	H	Q	T	P	W	Z	C	Y	E	F	J	N	L	R	U	X	A	D	G	K	Q
W	Y	F	J	S	P	E	N	Q	M	T	W	Z	V	B	C	G	K	I	O	R	U	X	A	D	H	L
X	V	C	G	P	M	B	K	N	J	Q	T	W	S	Y	Z	D	H	F	L	O	R	U	X	A	E	I
Y	R	Y	C	L	I	X	G	J	F	M	P	S	O	U	V	Z	D	B	H	K	N	Q	T	W	A	E
Z	N	U	Y	H	E	T	C	F	B	I	L	O	K	Q	R	V	Z	X	D	G	J	M	P	S	W	A

TABLE III

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/1}=\Theta_{1/2}$; $\Theta_{p/1}=\Theta_{c/2}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X
B	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O
C	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N
D	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W
E	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L
F	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A
G	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V
H	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K
I	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M
J	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U
K	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J
L	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D
M	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T
N	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H
O	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E
P	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S
Q	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G
R	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I
S	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z
T	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q
U	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C
V	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R
W	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y
X	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P
Y	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B
Z	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F

TABLE IV

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/1}=\theta_{i/2}$; $\theta_{p/2}=\theta_{c/1}$ ($\theta_{i/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	U	B	F	O	L	A	J	M	I	P	S	V	R	X	Y	C	G	E	K	N	Q	T	W	Z	D	H
B	V	C	G	P	M	B	K	N	J	Q	T	W	S	Y	Z	D	H	F	L	O	R	U	X	A	E	I
C	W	D	H	Q	N	C	L	O	K	R	U	X	T	Z	A	E	I	G	M	P	S	V	Y	B	F	J
D	X	E	I	R	O	D	M	P	L	S	V	Y	U	A	B	F	J	H	N	Q	T	W	Z	C	G	K
E	Y	F	J	S	P	E	N	Q	M	T	W	Z	V	B	C	G	K	I	O	R	U	X	A	D	H	L
F	Z	G	K	T	Q	F	O	R	N	U	X	A	W	C	D	H	L	J	P	S	V	Y	B	E	I	M
G	A	H	L	U	R	G	P	S	O	V	Y	B	X	D	E	I	M	K	Q	T	W	Z	C	F	J	N
H	B	I	M	V	S	H	Q	T	P	W	Z	C	Y	E	F	J	N	L	R	U	X	A	D	G	K	O
I	C	J	N	W	T	I	R	U	Q	X	A	D	Z	F	G	K	O	M	S	V	Y	B	E	H	L	P
J	D	K	O	X	U	J	S	V	R	Y	B	E	A	G	H	L	P	N	T	W	Z	C	F	I	M	Q
K	E	L	P	Y	V	K	T	W	S	Z	C	F	B	H	I	M	Q	O	U	X	A	D	G	J	N	R
L	F	M	Q	Z	W	L	U	X	T	A	D	G	C	I	J	N	R	P	V	Y	B	E	H	K	O	S
M	G	N	R	A	X	M	V	Y	U	B	E	H	D	J	K	O	S	Q	W	Z	C	F	I	L	P	T
N	H	O	S	B	Y	N	W	Z	V	C	F	I	E	K	L	P	T	R	X	A	D	G	J	M	Q	U
O	I	P	T	C	Z	O	X	A	W	D	G	J	F	L	M	Q	U	S	Y	B	E	H	K	N	R	V
P	J	Q	U	D	A	P	Y	B	X	E	H	K	G	M	N	R	V	T	Z	C	F	I	L	O	S	W
Q	K	R	V	E	B	Q	Z	C	Y	F	I	L	H	N	O	S	W	U	A	D	G	J	M	P	T	X
R	L	S	W	F	C	R	A	D	Z	G	J	M	I	O	P	T	X	V	B	E	H	K	N	Q	U	Y
S	M	T	X	G	D	S	B	E	A	H	K	N	J	P	Q	U	Y	W	C	F	I	L	O	R	V	Z
T	N	U	Y	H	E	T	C	F	B	I	L	O	K	Q	R	V	Z	X	D	G	J	M	P	S	W	A
U	O	V	Z	I	F	U	D	G	C	J	M	P	L	R	S	W	A	Y	E	H	K	N	Q	T	X	B
V	P	W	A	J	G	V	E	H	D	K	N	Q	M	S	T	X	B	Z	F	I	L	O	R	U	Y	C
W	Q	X	B	K	H	W	F	I	E	L	O	R	N	T	U	Y	C	A	G	J	M	P	S	V	Z	D
X	R	Y	C	L	I	X	G	J	F	M	P	S	O	U	V	Z	D	B	H	K	N	Q	T	W	A	E
Y	S	Z	D	M	J	Y	H	K	G	N	Q	T	P	V	W	A	E	C	I	L	O	R	U	X	B	F
Z	T	A	E	N	K	Z	I	L	H	O	R	U	Q	W	X	B	F	D	J	M	P	S	V	Y	C	G

TABLE V

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/2}=\theta_{p/1}$; $\theta_{1/1}=\theta_{c/2}$ ($\theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L
B	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P
C	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q
D	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J
E	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H
F	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B
G	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S
H	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T
I	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G
J	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U
K	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V
L	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W
M	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K
N	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O
O	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X
P	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y
Q	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z
R	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C
S	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E
T	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D
U	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M
V	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A
W	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N
X	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F
Y	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R
Z	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I

TABLE VI

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/2}=\theta_{c/1}$; $\theta_{1/1}=\theta_{p/2}$ ($\theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	T	P	G	J	U	L	I	M	F	C	Z	D	X	W	S	O	Q	K	H	E	B	Y	V	R	N
B	H	A	W	N	Q	B	S	P	T	M	J	G	K	E	D	Z	V	X	R	O	L	I	F	C	Y	U
C	L	E	A	R	U	F	W	T	X	Q	N	K	O	I	H	D	Z	B	V	S	P	M	J	G	C	Y
D	U	N	J	A	D	O	F	C	G	Z	W	T	X	R	Q	M	I	K	E	B	Y	V	S	P	L	H
E	R	K	G	X	A	L	C	Z	D	W	T	Q	U	O	N	J	F	H	B	Y	V	S	P	M	I	E
F	G	Z	V	M	P	A	R	O	S	L	I	F	J	D	C	Y	U	W	Q	N	K	H	E	B	X	T
G	P	I	E	V	Y	J	A	X	B	U	R	O	S	M	L	H	D	F	Z	W	T	Q	N	K	G	C
H	S	L	H	Y	B	M	D	A	E	X	U	R	V	P	O	K	G	I	C	Z	W	T	Q	N	J	F
I	O	H	D	U	X	I	Z	W	A	T	Q	N	R	L	K	G	C	E	Y	V	S	P	M	J	F	B
J	V	O	K	B	E	P	G	D	H	A	X	U	Y	S	R	N	J	L	F	C	Z	W	T	Q	M	I
K	Y	R	N	E	H	S	J	G	K	D	A	X	B	V	U	Q	M	O	I	F	C	Z	W	T	P	L
L	B	U	Q	H	K	V	M	J	N	G	D	A	E	Y	X	T	P	R	L	I	F	C	Z	W	S	O
M	X	Q	M	D	G	R	I	F	J	C	Z	W	A	U	T	P	L	N	H	E	B	Y	V	S	O	K
N	D	W	S	J	M	X	O	L	P	I	F	C	G	A	Z	V	R	T	N	K	H	E	B	Y	U	Q
O	E	X	T	K	N	Y	P	M	Q	J	G	D	H	B	A	W	S	U	O	L	I	F	C	Z	V	R
P	I	B	X	O	R	C	T	Q	U	N	K	H	L	F	E	A	W	Y	S	P	M	J	G	D	Z	V
Q	M	F	B	S	V	G	X	U	Y	R	O	L	P	J	I	E	A	C	W	T	Q	N	K	H	D	Z
R	K	D	Z	Q	T	E	V	S	W	P	M	J	N	H	G	C	Y	A	U	R	O	L	I	F	B	X
S	Q	J	F	W	Z	K	B	Y	C	V	S	P	T	N	M	I	E	G	A	X	U	R	O	L	H	D
T	T	M	I	Z	C	N	E	B	F	Y	V	S	W	Q	P	L	H	J	D	A	X	U	R	O	K	G
U	W	P	L	C	F	Q	H	E	I	B	Y	V	Z	T	S	O	K	M	G	D	A	X	U	R	N	J
V	Z	S	O	F	I	T	K	H	L	E	B	Y	C	W	V	R	N	P	J	G	D	A	X	U	Q	M
W	C	V	R	I	L	W	N	K	O	H	E	W	F	U	T	U	Q	S	M	J	G	D	V	X	T	P
X	F	Y	U	L	O	Z	Q	N	R	K	H	A	I	C	B	X	T	V	P	M	J	G	D	A	W	S
Y	J	C	Y	P	S	D	U	R	V	O	L	I	M	G	F	B	X	Z	T	Q	N	K	H	E	A	W
Z	N	G	C	T	W	H	Y	V	Z	S	P	M	Q	K	J	F	B	D	X	U	R	O	L	I	E	A

TABLE VII

Components:

(1)—A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2)—F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/2}=\theta_{p/1}$; $\theta_{1/2}=\theta_{c/1}$ ($\theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
B	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
C	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
D	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
E	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
F	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
G	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
H	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
I	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
J	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
K	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
L	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
M	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
N	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
O	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
P	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Q	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
R	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
T	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
U	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
V	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
W	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
X	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
Y	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Z	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S

TABLE VIII

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/2}=\Theta_{c/1}$; $\Theta_{1/2}=\Theta_{p/1}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
D	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
E	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
F	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
G	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
H	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
K	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
L	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
M	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
N	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
O	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
P	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
Q	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
R	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
S	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
T	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
U	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
V	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
Y	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
Z	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G

TABLE IX[2]

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/1}=\theta_{p/2}$; $\theta_{i/1}=\theta_{c/2}$ ($\theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	V	F	R	T	S	X	I	E	Z	D	M	A	U	W	N	B	C	Y	G	H	J	K	L	O	P	Q
C	K	X	Y	H	G	O	Z	S	Q	T	U	V	J	L	W	F	R	P	I	E	D	M	A	N	B	C
D	M	O	P	E	I	N	Q	G	C	H	J	K	D	A	L	X	Y	B	Z	S	T	U	V	W	F	R
E	U	N	B	S	Z	W	C	I	R	E	D	M	T	V	A	O	P	F	Q	G	H	J	K	L	X	Y
F	J	W	F	G	Q	L	R	Z	Y	S	T	U	H	K	V	N	B	X	C	I	E	D	M	A	O	P
G	D	L	X	I	C	A	Y	Q	P	G	H	J	E	M	K	W	F	O	R	Z	S	T	U	V	N	B
H	T	A	O	Z	R	V	P	C	B	I	E	D	S	U	M	L	X	N	Y	Q	G	H	J	K	W	F
I	H	V	N	Q	Y	K	B	R	F	Z	S	T	G	J	U	A	O	W	P	C	I	E	D	M	L	X
J	E	K	W	C	P	M	F	Y	X	Q	G	H	I	D	J	V	N	L	B	R	Z	S	T	U	A	O
K	S	M	L	R	B	U	X	P	O	C	I	E	Z	T	D	K	W	A	F	Y	Q	G	H	J	V	N
L	G	U	A	Y	F	J	O	B	N	R	Z	S	Q	H	T	M	L	V	X	P	C	I	E	D	K	W
M	I	J	V	P	X	D	N	F	W	Y	Q	G	C	E	H	U	A	K	O	B	R	Z	S	T	M	L
N	Z	D	K	B	O	T	W	X	L	P	C	I	R	S	E	J	V	M	N	F	Y	Q	G	H	U	A
O	Q	T	M	F	N	H	L	O	A	B	R	Z	Y	G	S	D	K	U	W	X	P	C	I	E	J	V
P	C	H	U	X	W	E	A	N	V	F	Y	Q	P	I	G	T	M	J	L	O	B	R	Z	S	D	K
Q	R	E	J	O	L	S	V	W	K	X	P	C	B	Z	I	H	U	D	A	N	F	Y	Q	G	T	M
R	Y	S	D	N	A	G	K	L	M	O	B	R	F	Q	Z	E	J	T	V	W	X	P	C	I	H	U
S	P	G	T	W	V	I	M	A	U	N	F	Y	X	C	Q	S	D	H	K	L	O	B	R	Z	E	J
T	B	I	H	L	K	Z	U	V	J	W	X	P	O	R	C	G	T	E	M	A	N	F	Y	Q	S	D
U	F	Z	E	A	M	Q	J	K	D	L	O	B	N	Y	R	I	H	S	U	V	W	X	P	C	G	T
V	X	Q	S	V	U	C	D	M	T	A	N	F	W	P	Y	Z	E	G	J	K	L	O	B	R	I	H
W	O	C	G	K	J	R	T	U	H	V	W	X	L	B	P	Q	S	I	D	M	A	N	F	Y	Z	E
X	N	R	I	M	D	Y	H	J	E	K	L	O	A	F	B	C	G	Z	T	U	V	W	X	P	Q	S
Y	W	Y	Z	U	T	P	E	D	S	M	A	N	V	X	F	R	I	Q	H	J	K	L	O	B	C	G
Z	L	P	Q	J	H	B	S	T	G	U	V	W	K	O	X	Y	Z	C	E	D	M	A	N	F	R	I

---

[2] An interesting fact about this case is that if the plain component is made identical with the cipher component (both being the sequence FBPY . . .), and if the enciphering equations are the same as for Table 1-B, then the resultant cipher square is identical with Table IX, except that the key letters at the left are in the order of the reversed mixed component, FXON . . . . In other words, the secondary cipher alphabets produced by the interaction of two identical mixed components are the same as those given by the interaction of a mixed component and the normal component.

TABLE X[3]

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/1}=\theta_{c/2}$; $\theta_{i/1}=\theta_{p/2}$ ($\theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	L	P	Q	J	H	B	S	T	G	U	V	W	K	O	X	Y	Z	C	E	D	M	A	N	F	R	I
C	W	Y	Z	U	T	P	E	D	S	M	A	N	V	X	F	R	I	Q	H	J	K	L	O	B	C	G
D	N	R	I	M	D	Y	H	J	E	K	L	O	A	F	B	C	G	Z	T	U	V	W	X	P	Q	S
E	O	C	G	K	J	R	T	U	H	V	W	X	L	B	P	Q	S	I	D	M	A	N	F	Y	Z	E
F	X	Q	S	V	U	C	D	M	T	A	N	F	W	P	Y	Z	E	G	J	K	L	O	B	R	I	H
G	F	Z	E	A	M	Q	J	K	D	L	O	B	N	Y	R	I	H	S	U	V	W	X	P	C	G	T
H	B	I	H	L	K	Z	U	V	J	W	X	P	O	R	C	G	T	E	M	A	N	F	Y	Q	S	D
I	P	G	T	W	V	I	M	A	U	N	F	Y	X	C	Q	S	D	H	K	L	O	B	R	Z	E	J
J	Y	S	D	N	A	G	K	L	M	O	B	R	F	Q	Z	E	J	T	V	W	X	P	C	I	H	U
K	R	E	J	O	L	S	V	W	K	X	P	C	B	Z	I	H	U	D	A	N	F	Y	Q	G	T	M
L	C	H	U	X	W	E	A	N	V	F	Y	Q	P	I	G	T	M	J	L	O	B	R	Z	S	D	K
M	Q	T	M	F	N	H	L	O	A	B	R	Z	Y	G	S	D	K	U	W	X	P	C	I	E	J	V
N	Z	D	K	B	O	T	W	X	L	P	C	I	R	S	E	J	V	M	N	F	Y	Q	G	H	U	A
O	I	J	V	P	X	D	N	F	W	Y	Q	G	C	E	H	U	A	K	O	B	R	Z	S	T	M	L
P	G	U	A	Y	F	J	O	B	N	R	Z	S	Q	H	T	M	L	V	X	P	C	I	E	D	K	W
Q	S	M	L	R	B	U	X	P	O	C	I	E	Z	T	D	K	W	A	F	Y	Q	G	H	J	V	N
R	E	K	W	C	P	M	F	Y	X	Q	G	H	I	D	J	V	N	L	B	R	Z	S	T	U	A	O
S	H	V	N	Q	Y	K	B	R	F	Z	S	T	G	J	U	A	O	W	P	C	I	E	D	M	L	X
T	T	A	O	Z	R	V	P	C	B	I	E	D	S	U	M	L	X	N	Y	Q	G	H	J	K	W	F
U	D	L	X	I	C	A	Y	Q	P	G	H	J	E	M	K	W	F	O	R	Z	S	T	U	V	N	B
V	J	W	F	G	Q	L	R	Z	Y	S	T	U	H	K	V	N	B	X	C	I	E	D	M	A	O	P
W	U	N	B	S	Z	W	C	I	R	E	D	M	T	V	A	O	P	F	Q	G	H	J	K	L	X	Y
X	M	O	P	E	I	N	Q	G	C	H	J	K	D	A	L	X	Y	B	Z	S	T	U	V	W	F	R
Y	K	X	Y	H	G	O	Z	S	Q	T	U	V	J	L	W	F	R	P	I	E	D	M	A	N	B	C
Z	V	F	R	T	S	X	I	E	Z	D	M	A	U	W	N	B	C	Y	G	H	J	K	L	O	P	Q

---

[3] Footnote 2 to Table IX, page 104, also applies to this table, except that the key letters at the left will follow the order of the direct mixed component.

TABLE XI

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/1}=\theta_{p/2}$; $\theta_{1/2}=\theta_{c/1}$ ($\theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	G	Z	V	M	P	A	R	O	S	L	I	F	J	D	C	Y	U	W	Q	N	K	H	E	B	X	T
B	H	A	W	N	Q	B	S	P	T	M	J	G	K	E	D	Z	V	X	R	O	L	I	F	C	Y	U
C	I	B	X	O	R	C	T	Q	U	N	K	H	L	F	E	A	W	Y	S	P	M	J	G	D	Z	V
D	J	C	Y	P	S	D	U	R	V	O	L	I	M	G	F	B	X	Z	T	Q	N	K	H	E	A	W
E	K	D	Z	Q	T	E	V	S	W	P	M	J	N	H	G	C	Y	A	U	R	O	L	I	F	B	X
F	L	E	A	R	U	F	W	T	X	Q	N	K	O	I	H	D	Z	B	V	S	P	M	J	G	C	Y
G	M	F	B	S	V	G	X	U	Y	R	O	L	P	J	I	E	A	C	W	T	Q	N	K	H	D	Z
H	N	G	C	T	W	H	Y	V	Z	S	P	M	Q	K	J	F	B	D	X	U	R	O	L	I	E	A
I	O	H	D	U	X	I	Z	W	A	T	Q	N	R	L	K	G	C	E	Y	V	S	P	M	J	F	B
J	P	I	E	V	Y	J	A	X	B	U	R	O	S	M	L	H	D	F	Z	W	T	Q	N	K	G	C
K	Q	J	F	W	Z	K	B	Y	C	V	S	P	T	N	M	I	E	G	A	X	U	R	O	L	H	D
L	R	K	G	X	A	L	C	Z	D	W	T	Q	U	O	N	J	F	H	B	Y	V	S	P	M	I	E
M	S	L	H	Y	B	M	D	A	E	X	U	R	V	P	O	K	G	I	C	Z	W	T	Q	N	J	F
N	T	M	I	Z	C	N	E	B	F	Y	V	S	W	Q	P	L	H	J	D	A	X	U	R	O	K	G
O	U	N	J	A	D	O	F	C	G	Z	W	T	X	R	Q	M	I	K	E	B	Y	V	S	P	L	H
P	V	O	K	B	E	P	G	D	H	A	X	U	Y	S	R	N	J	L	F	C	Z	W	T	Q	M	I
Q	W	P	L	C	F	Q	H	E	I	B	Y	V	Z	T	S	O	K	M	G	D	A	X	U	R	N	J
R	X	Q	M	D	G	R	I	F	J	C	Z	W	A	U	T	P	L	N	H	E	B	Y	V	S	O	K
S	Y	R	N	E	H	S	J	G	K	D	A	X	B	V	U	Q	M	O	I	F	C	Z	W	T	P	L
T	Z	S	O	F	I	T	K	H	L	E	B	Y	C	W	V	R	N	P	J	G	D	A	X	U	Q	M
U	A	T	P	G	J	U	L	I	M	F	C	Z	D	X	W	S	O	Q	K	H	E	B	Y	V	R	N
V	B	U	Q	H	K	V	M	J	N	G	D	A	E	Y	X	T	P	R	L	I	F	C	Z	W	S	O
W	C	V	R	I	L	W	N	K	O	H	E	B	F	Z	Y	U	Q	S	M	J	G	D	A	X	T	P
X	D	W	S	J	M	X	O	L	P	I	F	C	G	A	Z	V	R	T	N	K	H	E	B	Y	U	Q
Y	E	X	T	K	N	Y	P	M	Q	J	G	D	H	B	A	W	S	U	O	L	I	F	C	Z	V	R
Z	F	Y	U	L	O	Z	Q	N	R	K	H	E	I	C	B	X	T	V	P	M	J	G	D	A	W	S

TABLE XII

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/1}=\theta_{e/2}$; $\theta_{1/2}=\theta_{p/1}$ ($\theta_{1/2}$ is F).

	PLAIN TEXT																									
KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B
B	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P
C	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y
D	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R
E	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C
F	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q
G	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z
H	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I
I	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G
J	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S
K	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E
L	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H
M	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T
N	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D
O	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J
P	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U
Q	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M
R	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K
S	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V
T	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A
U	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L
V	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W
W	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N
X	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O
Y	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X
Z	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F

APPENDIX 2[1]

ELEMENTARY STATISTICAL THEORY APPLICABLE TO THE PHENOMENA OF REPETITION IN CRYPTANALYSIS

1. Introductory.—*a*. In Par. 9*c* it was stated that the phenomena of repetition in cryptanalytics may be removed from the realm of intuition and dealt with statistically. The discussion of the matter will here be confined to relatively simple phases of the theory of probability, a definition of which implies philosophical questions of no practical interest to the student of cryptanalysis. For his purposes, the following definition of *a priori* probability will be sufficient:

> The probability that an event will occur is the ratio of the number of "favorable cases" to the number of total possible cases, all cases being equally likely to occur. By a "favorable case" is meant one which will produce the event in question.

b. In what follows, reference will be made to *random assortments* of letters and especially to *random text*. By the latter will be meant merely that the text under consideration has been assumed to have been enciphered by some more or less complex cryptographic system so that for all practical purposes the sequence of letters constituting this text is a random assortment; that is, the sequence is just about what would have been obtained if the letters had been drawn at random out of a box containing a large number of the 26 letters of the alphabet, all in equal proportions, so that there are exactly the same numbers of A's, B's, C's, . . . Z's. It is assumed that each time in making a drawing from such a box, the latter is thoroughly shaken so that the letters are thoroughly mixed and then a single letter is selected at random, recorded, and replaced in the same box. In what follows, the word "box" will refer to the box as described.

c. A uniliteral frequency distribution of a large volume of random text will be "flat," i. e., lacking crests and troughs.

d. For purposes of statistical analysis, the text of a monoalphabetic substitution cipher is equivalent to plain text. As a corollary, when a polyalphabetic substitution cipher has been reduced to the simple terms of a set of monoalphabets, i. e., when the letters constituting the cipher text have been allocated into their proper uniliteral distributions, the letters falling into the respective distributions are statistically equivalent to plain text.

2. Data pertaining to single letters.—*a*. (1) A single letter will be drawn at random from the box. What is the probability that it will be an A? According to the foregoing definition of probability, since the total number of possible cases is 26 and the number of favorable cases is here only 1, the probability is $1:26=\frac{1}{26}=.0385$. This is the probability of drawing an A from the box. The probability that the letter drawn will be a B, a C, a D, . . ., a Z is the same as for A. In other words, the probability of drawing *any specified single letter* is $p=.0385$.

(2) The value $p=.0385$, as found above, may also be termed the probability constant for single letters in random text of a 26-letter alphabet. For any language this constant is merely the reciprocal of the total number of different characters which may be employed in writing the text in question.

---

[1] In the preparation of this appendix, the author has had the benefit of the very helpful suggestions of Capt. H. G. Miller, Signal Corps, Mr. F. B. Rowlett, Dr. S. Kullback, and Dr. A. Sinkov, Assistant Cryptanalysts, O. C. Sig. O. Certain parts of Dr. Kullback's important paper "Statistical Methods in Cryptanalysis" form the basis of the discussion.

(3) Another way of interpreting the notation $p=.0385$ is to say that in a large volume of random text, for example in 100,000 letters, any letter that one may choose to specify may be expected to occur about 3,850 times; in 10,000 letters it may be expected to occur about 385 times; in 1,000 letters, about 38.5 times, and so on. In every-day language it would be said that "in the long run" or "on the average" in 1,000 letters of random text there will be about 38.5 occurrences of each of the 26 letters of the alphabet.

(4) But unfortunately, in cryptanalysis it is not often the case that one has such a large number of letters available for study in any single cipher alphabet. More often the cryptanalyst has a relatively small number of letters and these must be distributed over several cipher alphabets. Hence it is necessary to be able to deal with smaller numbers of letters. Consider a specific piece of random text of only 100 letters. It has been seen that "in the long run" each letter may be expected to occur about 3.85 times in this amount of random text; that is, the 26 letters will have an *average* frequency of 3.85. But in reaching this average of 3.85 occurrences in 100 letters, it is obvious that some letter or letters may not appear at all, some may appear once, some twice, and so on. How many will not appear at all; how many will appear 1, 2, 3, . . . times? In other words, how will the different categories of letters (different in respect to frequency of occurrence) be distributed, or what will the *distribution* be like? Will it follow any kind of law or pattern? The cryptanalyst also wants to know the answer to questions such as these: What is the probability that a specified letter will not appear at all in a given piece of text? That it will appear *exactly* 1, 2, 3, . . . times? That it will appear *at least* 1, 2, 3, . . . times? The same sort of questions may be asked with respect to digraphs, trigraphs, and so on.

b. (1) It may be stated at once that questions of this nature are not easily answered, and a complete discussion falls quite outside the scope of this text. However, it will be sufficient for the present purposes if the student is provided with a more or less simple and practical means of finding the answers. With this in view certain curves have been prepared from data based upon Poisson's exponential expansion, or the "law of small probabilities" and their use will now be explained. Students without a knowledge of the mathematical theory of probability and statistics will have to take the curves "on faith" Those interested in their derivation are referred to the following texts:

Fisher, R. A., *Statistical Methods for Research Workers*, London, 1937.
Fry, T. C., *Probability and Its Engineering Uses*, New York, 1928.

(2) By means of these *probability curves*, it is possible to find, in a relatively easy manner, the probability for 0, 1, 2, . . . 11 occurrences of an event in n cases, if the *mean* (expected, average, probable) number of occurrences in these n cases is known. For example, given a cryptogram equivalent to 100 letters of random text, what is the probability that any specified single letter, whatever will not appear at all in the cryptogram? Since the probability of the occurrence of a specified single letter is $\frac{1}{26}=.0385$, and there are 100 letters in the cryptogram, the average or expected or mean number of occurrences of an A, a B, a C, . . ., is $.0385\times100=3.85$. Refer now to that probability curve which is marked "f_0", meaning "frequency zero", or "zero occurrences." On the horizontal or x axis of that curve find the point corresponding to the value 3.85 and follow the vertical coordinate determined by this value up to the point of intersection with the curve itself; then follow the horizontal coordinate determined by this intersection point over to the left and read the value on the vertical axis of the curve. It is approximately .021. This means that the probability that a specified single letter (an A, a B, a C, . . .) will not appear at all in the cryptogram, if it really were a perfectly random assortment of 100 letters, is .021.

That is, according to the theory of probability, in 1,000 cases of random-text messages of 100 letters each, one may expect to find about 21 messages in which a specified single letter will not appear at all. Another way of saying the same thing is: If 1,000 sets of 100 letters of random text are examined, in about 21 out of the 1,000 such sets any letter that one may choose to name will be absent. This, of course, is merely a theoretical expectancy; it indicates only what probably will happen in the long run.

(3) What is the probability that a specified single letter will appear *exactly* once in 100 letters of random text? To answer this question, find on the curve marked f_1, the point of intersection of the vertical coordinate corresponding to the mean or average value 3.85 with the curve; follow the horizontal coordinate thus determined over to the vertical scale at the left; read the value on this scale. It is .082, which means that in 1,000 cases of random-text messages of 100 letters each, one may expect to find about 82 messages in which any letter one chooses to specify will occur exactly once, no more and no less.

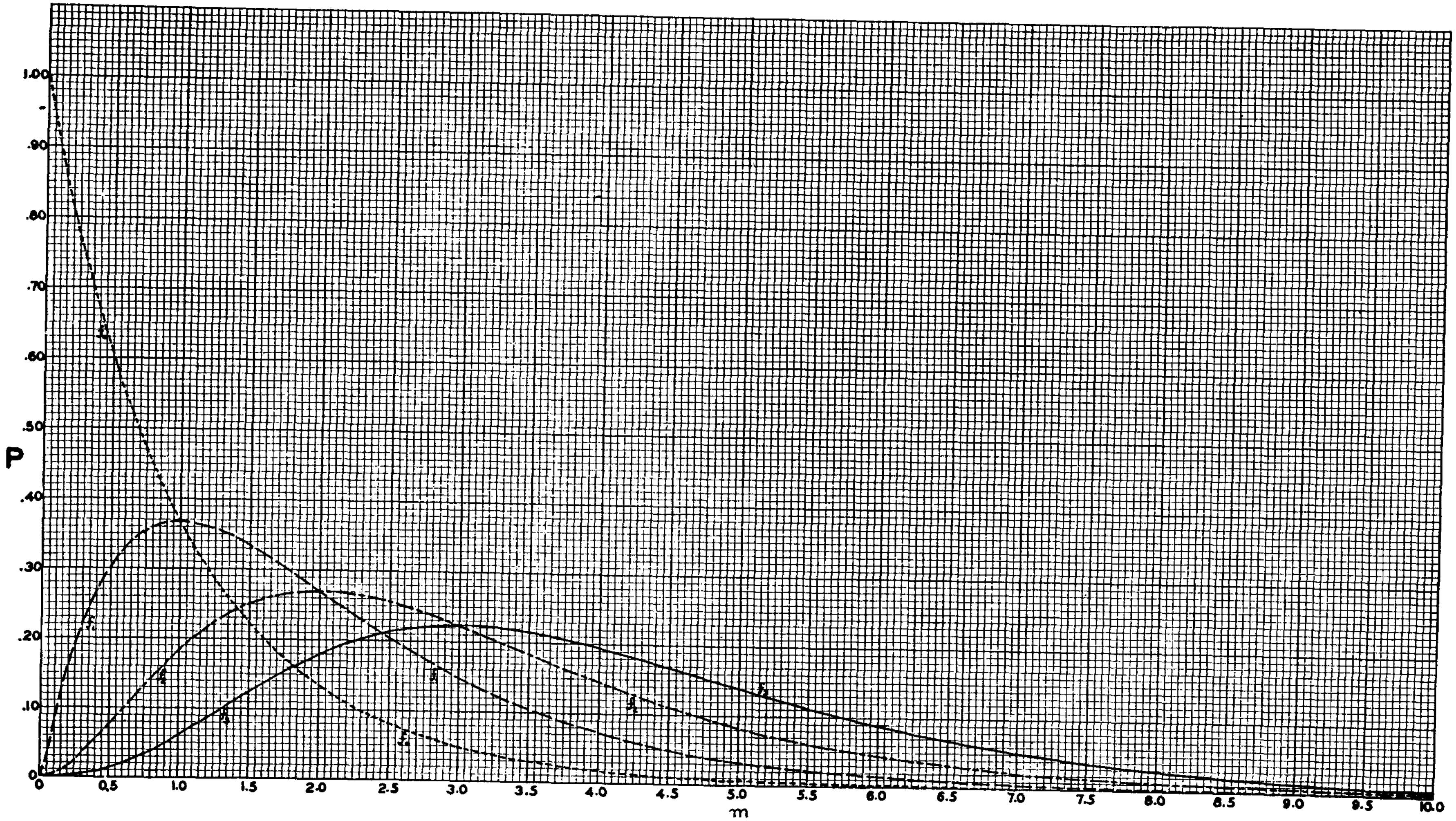
(4) In the same way, the probability that a specified single letter will appear *exactly* twice is found to be .158; exactly 3 times, .202; and so on, as shown in the table below:

100 letters of random text

Frequency (x)	Probability that a specified single letter will occur exactly x times
0	0. 021
1	. 082
2	. 158
3	. 202
4	. 195
5	. 150
6	. 096
7	. 053
8	. 026
9	. 011
10	. 004
11	. 001

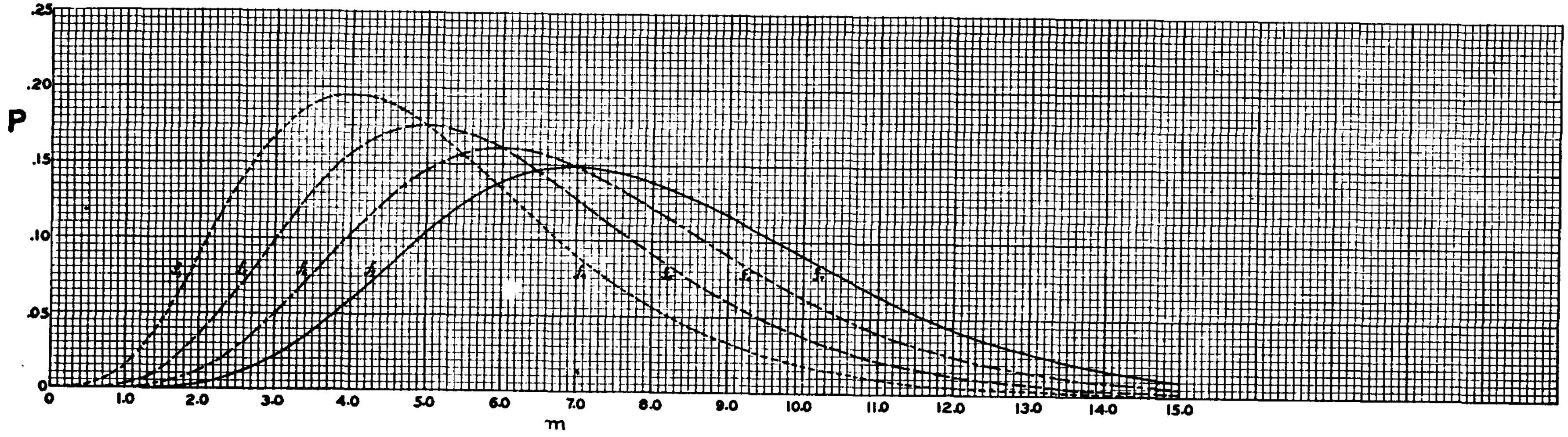
(5) To find the probability that a specified single letter will occur *at least* 1, 2, 3, . . . times in a series of letters constituting random text, one reasons as follows: Since the concept "at least 1" implies that the number specified is to be considered only as the minimum, with no limit indicated as to maximum, occurrences of 2, 3, 4, . . . are also "favorable" cases; the probabilities for *exactly* 1, 2, 3, 4, . . . occurrences should therefore be added and this will give the probability for "at least 1." Thus, in the case of 100 letters, the sum of the probabilities for exactly 1 to 11 occurrences, as set forth in the table directly above, is .978, and the latter value approximates the probability for at least 1 occurrence.

(6) A more accurate result will be obtained by the following reasoning. The probability for zero occurrences is .021. Since it is certain that a specified letter will occur either zero times or 1, 2, 3, . . . times, to find the probability for *at least* one time it is merely necessary to subtract the probability for zero occurrences from unity. That is, $1-.021=.979$, which is .001 greater than the result obtained by the other method. The reason it is greater is that the value .979 includes occurrences beyond 11, which were excluded from the previous calculation. Of course, the probabilities for these occurrences beyond 11 are very small, but taken all together they

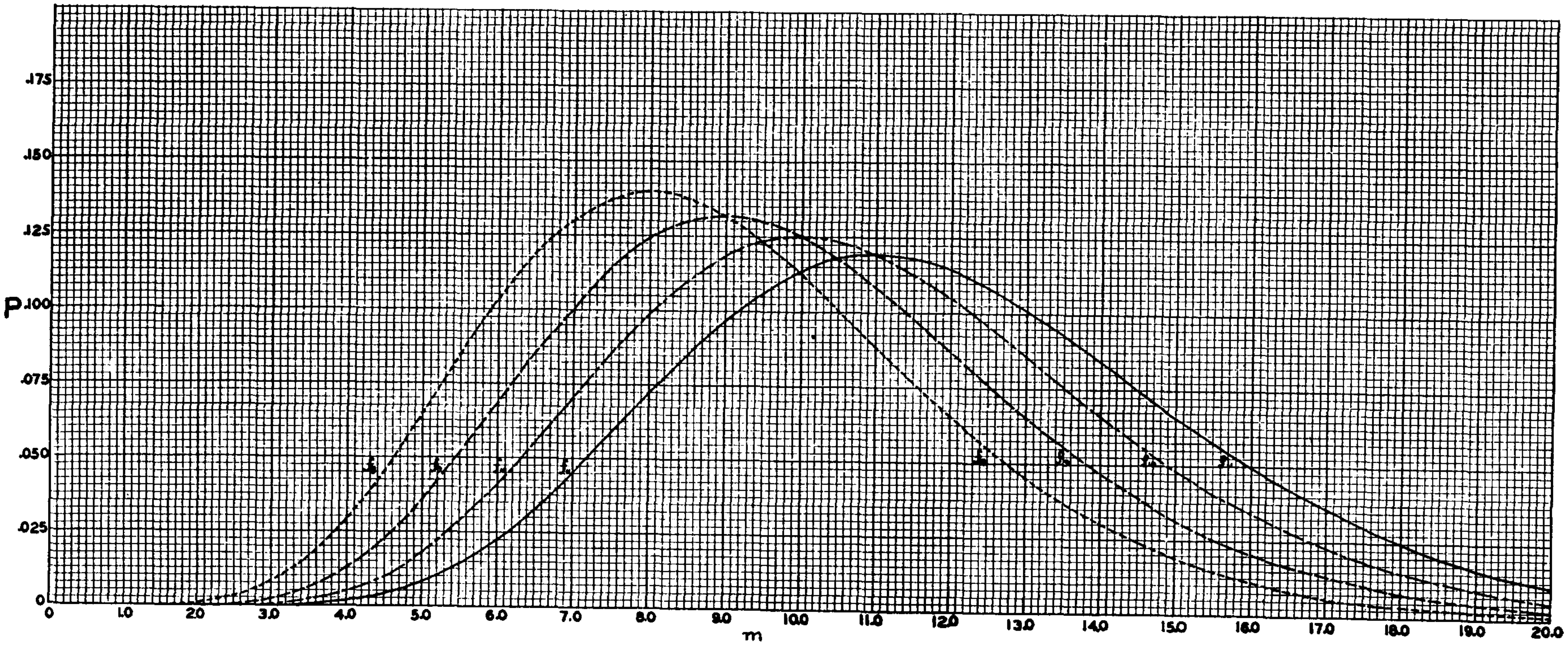


Curves showing probability for 0, 1, 2, and 3 occurrences of an event in n cases, given the mean number of occurrences.

(Face p. 110) No. 1



Curves showing probability for 4, 5, 6, and 7 occurrences of an event in *n* cases, given the mean number of occurrences.



Curves showing probability for 8, 9, 10, and 11 occurrences of an event in *n* cases, given the mean number of occurrences.

(Face p. 110) No. 2

add up to .001, the difference between the results obtained by the two methods. The probability for at least 2 occurrences is the difference between unity and the sum of the probability for zero and exactly 1 occurrences; that is, $1-(P_0+P_1)=1-(.021+.082)=1-.103=.897$. The respective probabilities for various numbers of occurrences of a specified single letter (from 0 to 11) are given in the following table:

100 letters of random text

Frequency (x)	Probability that a specified single letter will occur exactly x times	Probability that a specified single letter will occur at least x times
0	0. 021	1. 000
1	. 082	. 979
2	. 158	. 897
3	. 202	. 739
4	. 195	. 537
5	. 150	. 342
6	. 096	. 192
7	. 053	. 096
8	. 026	. 043
9	. 011	. 017
10	. 004	. 006
11	. 001	. 002

(7) The foregoing calculations refer to random text composed of 100 letters. For other numbers of letters, it is merely necessary to find the mean (multiply the probability for drawing a specified single letter out of the box, which is $\frac{1}{26}$ or .0385, by the number of letters in the assortment) and refer to the various curves, as before. For example, for a random assortment of 200 letters, the mean is $200 \times .0385$, or 7.7, and this is the value of the point to be sought along the horizontal or x axes of the curves; the intersections of the respective vertical lines corresponding to this mean with the various curves for 0, 1, 2, 3, . . . occurrences give the probabilities for these occurrences, the reading being taken on the vertical or y axes of the curves.

(8) The discussion thus far has dealt with the probabilities for 0, 1, 2, 3, . . . occurrences of specified single letters. It may be of more practical advantage to the student if he could be shown how to find the answer to these questions: Given a random assortment of 100 letters *how many* letters may be expected to occur *exactly* 0, 1, 2, 3, . . . times? How many may be expected to occur *at least* 1, 2, 3, . . . times? The curves may here again be used to answer these questions, by a very simple calculation: multiply the probability value as obtained above for a specified single letter by the number of different elements being considered. For example, the probability that a specified single letter will occur exactly twice in a perfectly random assortment of 100 letters is .158; since the number of different letters is 26, the absolute number of single letters that may be expected to occur exactly 2 times in this assortment is $.158 \times 26=4.108$. That is, in 100 letters of random text there should be about four letters which occur exactly 2 times. The following table gives the data for various numbers of occurrences.

100 letters of random text

Frequency (x)	Probability that a specified single letter will occur exactly x times	Probability that a specified single letter will occur at least x times	Probable number of letters appearing exactly x times	Probable number of letters appearing at least x times
0	0.021	1.000	0.546	26.000
1	.082	.979	2.132	25.454
2	.158	.897	4.108	23.322
3	.202	.739	5.252	19.214
4	.195	.537	5.070	13.962
5	.150	.342	3.900	8.892
6	.096	.192	2.496	4.992
7	.053	.096	1.378	2.496
8	.026	.043	.676	1.118
9	.011	.017	.286	.442
10	.004	.006	.104	.156
11	.001	.002	.026	.052

(9) Referring again to the curves, and specifically to the tabulated results set forth directly above, it will be seen that the probability that there will be exactly two occurrences of a specified single letter in 100 letters of random text (.158), is less than the probability that there will be exactly three occurrences (.202); in other words, the chances that a specified single letter will occur exactly three times are better, by about 25 percent, than that it will occur only two times. Furthermore, there will be about five letters which will occur exactly 3 times, and about five which will occur exactly 4 times, whereas there will be only about two letters which will occur exactly 1 time. Other facts of a similar import may be deduced from the foregoing table.

c. The discussion thus far has dealt with random assortments of letters. What about other types of texts, for example, normal plain text? What is the probability that E will occur 0, 1, 2, 3, . . . times in 50 letters of normal English? The relative frequency value or probability that a letter selected at random from a large volume of normal English text will be E is .12604. (In 100,000 letters E occurred 12,604 times.) For 50 letters this value must be multiplied by 50, giving 6.3 as the mean or point to be found along the x axes of the curves. The probabilities for 0, 1, 2, 3, . . . occurrences are tabulated below:

50 letters of normal English plain text

Frequency (x)	Probability that an E will be drawn exactly x times	Probability that an E will be drawn at least x times
0	0.002	1.000
1	.011	.998
2	.036	.987
3	.076	.951
4	.120	.875
5	.151	.755
6	.159	.604
7	.143	.445
8	.113	.302
9	.079	.223
10	.050	.173
11	.029	.123

d. (1) It has been seen that the probability of occurrence of a specified single letter in random text employing a 26-letter alphabet is $p=\frac{1}{26}=.0385$. If a considerable volume of such text is written on a large sheet of paper and a pencil is directed at random toward this text, the probability that the pencil point will hit the letter A, *or any other letter which may be specified in advance,* is .0385. Now suppose two pencils are directed simultaneously toward the sheet of paper. The probability that both pencil points will hit two A's is $\frac{1}{26}\times\frac{1}{26}=\frac{1}{26^2}=.00148$, since in this case one is dealing with the probability of the simultaneous occurrence of two events which are independent. The probability of hitting two B's, two C's, . . ., two Z's is likewise $\frac{1}{26^2}$. Hence, if no particular letter is specified, and merely this question is asked: "What is the probability that both pencil points will hit the same letter?" the answer must be the sum of the separate probabilities for simultaneously hitting two A's, two B's, and so on, for the whole alphabet, which is $26\times\frac{1}{26^2}=\frac{1}{26}=.0385$. This, then, is the probability that *any* two letters selected *at random* in random text of a 26-letter alphabet will be identical or will *coincide*. Since this value remains the same so long as the number of alphabetic elements remains fixed, it may be said that *the probability of monographic coincidence in random text of a 26-element alphabet is* .0385. The foregoing italicized expression [2] is important enough to warrant assigning a special symbol to it, *viz,* κ_r (read "kappa sub-*r*"). For a 26-element alphabet, then, $\kappa_r=.0385$.

(2) Now if one asks: "Given a random assortment of 10 letters, what are the respective probabilities of occurrence of 0, 1, 2, . . . single-letter coincidences?" one proceeds as follows. As before, it is first necessary to find the mean or expected number of coincidences and then refer to the various probability curves. To find the mean, one reasons as follows. Given a sequence of 10 letters, one may begin with the 1st letter and compare it with the 2d, 3d, . . . 10th letter to see if any two letters coincide; 9 such comparisons may be made, or in other words there are, beginning with the 1st letter, 9 opportunities for the occurrence of a coincidence. But one may also start with the 2nd letter and compare it with the 3d, 4th . . . 10th letter, thus yielding 8 more opportunities for the occurrence of a coincidence, and so on. This process may continue until one reaches the 9th letter and compares it with the 10th, yielding but one opportunity for the occurrence in question. The total number of comparisons that can be made is therefore the sum of the series of numbers 9, 8, 7, . . . 1, which is 45 comparisons.[3] Since in the 10 letters there are 45 opportunities for coincidence of single letters, and since the probability

---

[2] The expression itself may be termed a *parameter*, which in mathematics is often used to designate a constant that characterizes by each of its particular values some particular member of a system of values, functions, etc. The word is applicable in the case under discussion because the value obtained for κ_r is .0385; for a 25-element alphabet, $\kappa_r=.0400$; for a 27-element alphabet, $\kappa_r=.0370$, etc.

[3] The number of comparisons may readily be found by the formula $\frac{n(n-1)}{2}$, where n is the total number of letters involved. This formula is merely a special case under the general formula for ascertaining the number of combinations that may be made of n different things taken r at a time, which is ${}_nC_r=\frac{n!}{r!(n-r)!}$. In the present case, since only two letters are compared at a time, r is always 2, and hence the expression $\frac{n!}{r!(n-r)!}$, which is the same as $\frac{n(n-1)(n-2)!}{2(n-2)!}$, becomes by cancellation of the term $(n-2)!$ reduced to $\frac{n(n-1)}{2}$.

for monographic coincidence in random text is .0385 the expected number of coincidences is $.0385 \times 45 = 1.7325$. With $m=1.7$ one consults the various probability curves and an approximate distribution for exactly and for at least 0, 1, 2, . . . coincidences may readily be ascertained.[4]

e. (1) Now consider the matter of monographic coincidence in English plain text.[5] Following the same reasoning outlined in subpar. *d* (1), the probability of coincidence of two A's in plain text is the square of the probability of occurrence of the single letter A in such text. The probability of coincidence of two B's is the square of the probability of occurrence of the single letter B, and so on. The sum of these squares for all the letters of the alphabet, as shown in the following table, is found to be .0667.

Letter	Frequency [1] in 1,000 letters	Probability of separate occurrence of the letter	Square of probability of separate occurrence
A	73.66	0.0737	0.0054
B	9.74	.0097	.0001
C	30.68	.0307	.0009
D	42.44	.0424	.0018
E	129.96	.1300	.0169
F	28.32	.0283	.0008
G	16.38	.0164	.0003
H	33.88	.0339	.0012
I	73.52	.0735	.0054
J	1.64	.0016	.0000
K	2.96	.0030	.0000
L	36.42	.0364	.0013
M	24.74	.0247	.0006
N	79.50	.0795	.0063
O	75.28	.0753	.0057
P	26.70	.0267	.0007
Q	3.50	.0035	.0000
R	75.76	.0758	.0057
S	61.16	.0612	.0037
T	91.90	.0919	.0084
U	26.00	.0260	.0007
V	15.32	.0153	.0002
W	15.60	.0156	.0002
X	4.62	.0046	.0000
Y	19.34	.0193	.0004
Z	.98	.0010	.0000
Total	1,000.00	1.0000	.0667

[1] The data given are taken from Table 3, Appendix 1, Military Cryptanalysis, Part I.

This then is the probability that any two letters selected at random in a large volume of normal English telegraphic plain text will coincide. Since this value remains the same so long as the character of the language does not change radically, it may be said that *the probability of monographic coincidence in English telegraphic plain text is* .0667, or $\kappa_p = .0667$.

---

[4] The approximation given by the Poisson distribution in the case of single letters is not as good as that in the case of digraphs, trigraphs, etc., discussed in paragraphs 3, 4, below.

[5] The theory of monographic coincidence in plain text was originally developed and applied by the author in a technical paper written in 1925 dealing with his solution of messages enciphered by a cryptograph known as the "Hebern Electric Super-Code." The paper was printed in 1934.

(2) Given 10 letters of English plain text, what is the probability that there will be 0, 1, 2, . . . single-letter coincidences? Following the line of reasoning in subparagraph *d* (2), the expected number of coincidences is .0667×45=3.00, or $m=3$. The distribution for exactly and for at least 0, 1, 2, . . . coincidences may readily be found by reference to the various probability curves. (See footnote 4.)

f. The fact that κ_p (for English) is almost twice as great as κ_r is of considerable importance in cryptanalysis. It will be dealt with in detail in a subsequent text. At this point it will merely be said that κ_p and κ_r for other languages and alphabets have been calculated and show considerable variation, as will be noted in the table shown in paragraph 3*d*.

3. Data pertaining to digraphs.—*a.* (1) The foregoing discussion has been restricted to questions concerning single letters, but by slight modification it can be applied to questions concerning digraphs, trigraphs, and longer polygraphs.

(2) In the preceding cases it was necessary, before referring to the various probability curves, to find the mean or expected number of occurrences of the event in question in the total number of cases or trials being considered. Given a piece of random text totalling 100 letters, for example, what is the mean (average, probable, expected) number of occurrences of digraphs in this text? Since there are 676 different digraphs, the probability of occurrence of any specified digraph is $\frac{1}{676}=.00148$; since in 100 letters there are 99 digraphs (if the letters are taken consecutively in pairs) the mean or average number of occurrences in this case is .00148×99=.147. Having the mean number of occurrences of the event under consideration, one may now find the answers to these questions: What is the probability that any specified digraph, say XY, will not occur? What is the probability that it will occur *exactly* 1, 2, 3, . . . times? *At least* 1, 2, 3, . . . times?

(3) Again the probability curves may be used as before, for the type of distribution is the same. The following values are obtainable by reference to the various curves, using the mean value .00148×99=.147.

100 letters of random text

Frequency (x)	Probability that a specified digraph will occur exactly x times	Probability that a specified digraph will occur at least x times	Probable number of digraphs appearing exactly x times	Probable number of digraphs appearing at least x times
0	0. 86	1. 00	581. 36	676. 00
1	. 13	. 14	87. 88	94. 64
2	. 01	. 01	6. 76	6. 76
3	. 00	. 00	0. 00	0. 00

(4) Thus it is seen that in 100 letters of random text the probability that a specified digraph will occur exactly once, for example, is .13; at least once, .14; at least twice, .01. The probability that a specified digraph will occur at least 3 times is negligible. (By calculation, it is found to to be .0005.)

b. (1) The probability of digraphic coincidence in random text based upon a 26-element alphabet is of course quite simply obtained: since there are 26^2 different digraphs, the probability of selecting any specified digraph in random text is $\frac{1}{26^2}\cdot$ The probability of selecting two identical digraphs in such text, *when the digraphs are specified,* is $\frac{1}{26^2}\times\frac{1}{26^2}=\frac{1}{26^4}\cdot$ Since there are 26^2 different digraphs, the probability of digraphic coincidence in random text, $\kappa_{r}{}^{2}$, is $26^2\times\frac{1}{26^4}=\frac{1}{26^2}=$ .00148.

(2) Given a random assortment of 100 letters, what is the probability of occurrence of 0, 1, 2, . . . digraphic coincidences? Following the line of reasoning in paragraph 2*d* (2), in 100 letters the total number of comparisons that may be made to see if two digraphs coincide is 4,851. This number is obtained as follows: Consider the 1st and 2d letters in the series of 100 letters; they may be combined to form a digraph to be compared with the digraphs formed by combining the 2d and 3d, the 3d and 4th, the 4th and 5th letters, and so on, giving a total of 98 comparisons. Consider the digraph formed by combining the 2d and 3d letters; it may be compared with the digraphs formed by combining the 3d and 4th, 4th and 5th letters, and so on, giving a total of 97 comparisons. This process may be continued down to the digraph formed by combining the 98th and 99th letters, which yields only one comparison, since it may be compared only with the digraph resulting from combining the 99th and 100th letters. The total number of comparisons is the sum of the sequence of numbers 98, 97, 96, 95, . . . 1, which is 4,851.[6]

(3) Since in the 100 letters there are 4,851 opportunities for the occurrence of a digraphic coincidence, and since $\kappa_r{}^2=.00148$, the expected number of coincidences is $.00148\times4851=7.17948=7.2$. The various probability curves may now be referred to and the following results are obtained:

Distribution for 100 letters of random text

Frequency (*x*)	Probability for exactly *x* digraphic coincidences	Probability for at least *x* digraphic coincidences
0	0.001	1.000
1	.005	.999
2	.019	.994
3	.046	.975
4	.083	.929
5	.120	.846
6	.144	.726
7	.148	.582
8	.134	.434
9	.107	.300
10	.077	.193
11	.050	.116

c. In this table it will be noted that it is almost certain that in 100 letters of random text there will be at least one digraphic coincidence, despite the fact that there are 676 possible digraphs and only 99 of them have appeared in 100 letters. When one thinks of a total of 676 different digraphs from which the 99 digraphs may be selected it may appear rather incredible that the chances are better than even (.582) that one will find at least 7 digraphic coincidences in 100 letters of random text, yet that is what the statistical analysis of the problem shows to be the case. *These are, of course, purely accidental repetitions.* It is important that the student should fully realize that more coincidences or accidental repetitions than he feels intuitively should occur in random text will actually occur in the cryptograms he will study. He must therefore be on guard against putting too much reliance upon the surface appearances of the phenomena of repetition; he must calculate what may be expected from pure chance, to make sure that the number and length of the repetitions he does see in a cryptogram are really better than what may be expected in random text. In studying cryptograms composed of figures this

---

[6] The formula for finding the number of comparisons that can be made is as follows, where n = the total number of letters in the sequence and t is the length of the polygraph: Since the number of polygraphs possible is $n-t+1$, the number of comparisons is

$$\frac{(n-t+1)\ (n-t)}{2}$$

because any one of the $n-t+1$ polygraphs may be compared with any one of the remaining $n-t$ but as a comparison of A with B is the same as a comparison of B with A, the product must be halved.

is very important, for as the number of different symbols decreases the probability for purely chance coincidences increases.

d. (1) For convenience the following values of the reciprocals of various numbers from 20 to 36, and of the reciprocals of the squares, cubes, and 4th powers of these numbers are listed:

x	$1/x$	$1/x^2$	$1/x^3$	$1/x^4$
20	0. 0500	0. 002500	0. 000125	0. 00000625
21	. 0476	. 002266	. 000108	. 00000514
22	. 0455	. 002070	. 000094	. 00000429
23	. 0435	. 001892	. 000082	. 00000358
24	. 0417	. 001739	. 000073	. 00000302
25	. 0400	. 001600	. 000064	. 00000256
26	. 0385	. 001482	. 000057	. 00000220
27	. 0370	. 001369	. 000051	. 00000187
28	. 0357	. 001274	. 000046	. 00000162
29	. 0345	. 001190	. 000041	. 00000142
30	. 0333	. 001109	. 000037	. 00000123
31	. 0323	. 001043	. 000034	. 00000109
32	. 0313	. 000980	. 000031	. 00000096
33	. 0303	. 000918	. 000028	. 00000084
34	. 0294	. 000864	. 000025	. 00000075
35	. 0286	. 000818	. 000023	. 00000067
36	. 0278	. 000773	. 000021	. 00000060

(2) The following table gives the probabilities for monographic and digraphic coincidence for plain-text in several languages.

Language	κ_p	κ_{p^2}
English	0. 0667	0. 0069
French	. 0778	. 0093
German	. 0762	. 0112
Italian	. 0738	. 0081
Spanish	. 0775	. 0093

4. **Data pertaining to trigraphs, etc.**—*a.* Enough has been shown to make clear to the student how to calculate probability data concerning trigraphs, tetragraphs, and longer polygraphs.

b. (1) For example, in 100 letters of random text the value of m (the mean) for trigraphs is $.00005689\times100=.005689$. With so small a value, the probability curves are hardly usable, but at any rate they show that the probability of occurrence of a specified trigraph in so small a volume of text is so small as to be practically negligible. The probability of a specified trigraph occurring twice in that text is an even smaller quantity.

(2) The calculation for finding the probability of at least one trigraphic coincidence in 100 letters of random text is as follows:

$$m=\left(\frac{97\times98}{2}\right)\left(\frac{1}{26^3}\right)=4{,}753\times.0000568912=.2704=.27$$

Referring to curve f_0, with $m=.27$ the probability of finding no trigraphic coincidence is .76. The probability of finding at least one trigraphic coincidence is therefore $1-.76=.24$.

c. The calculation for a tetragraphic coincidence is as follows:

$$m=\left(\frac{96\times97}{2}\right)\left(\frac{1}{26^4}\right)=4{,}656\times.0000021883=.0101=.01$$

Referring to curve f_0, with $m=.01$ the probability of finding no tetragraphic coincidence is so high as to amount almost to certainty. Consequently, the probability of finding at least one tetragraphic coincidence is practically nil. (It is calculated to be .0094=approximately .01. This means that in a hundred cases of 100-letter random-text cryptograms, one might expect to find but one cryptogram in which a 4-letter repetition is brought about purely by chance; it is, in common parlance, a "hundred to one shot.") Consequently, if a tetragraphic repetition is found in a cryptogram of 100 letters, the probability that it is an accidental repetition is extremely small. If not accidental, then it must be causal, and the cause should be ascertained.

5. An example.—*a.* The message of Par. 9*a* of the text proper will be employed. First, let the repetitions be sought and underlined; then the repetitions are listed for convenience.

A. USYES ECPMP LCCLN XBWCS OXUVD
B. SCRHT HXIPL IBCIJ USYEE GURDP
C. AYBCX OFPJW JEMGP XVEUE LEJYQ
D. MUSCX JYMSG LLETA LEDEC GBMFI

Group	Number of occurrences
BC	2
CX	2
EC	2
LE	3
JY	2
PL	2
SC	2
SY	2
US	3
YE	2
SYE	2
USY	2
USYE	2

b. Referring to the table in Par. 3*a* (3) above, it will be seen that in 100 letters of random text one might expect to find about 7 digraphs appearing at least twice and no digraph appearing 3 times. The list of repetitions shows 8 digraphs occurring twice and 2 occurring 3 times.

c. Again, the list of repetitions shows 10 digraphs each repeated at least twice; the table in Par. 3*b* (3) above shows that in 100 letters of random text the probability of finding at least that many digraphic coincidences is only .193. That is, the chances of this being an accident are but 176 in a thousand; or another way of expressing the same thing is to say that the odds against this phenomenon being an accident are as 807 is to 193 or roughly 4 to 1.

d. The probability of finding at least one trigraphic coincidence in 100 letters of random text is very small, as noted in Par. 4*b;* the probability of finding at least one tetragraphic coincidence is still smaller (Par. 4*c*). Yet this cipher message of but 100 letters contains a repetition of this length.

e. A consideration of the foregoing leads to the conclusion that the number and length of the repetitions manifested by the cryptogram are not accidental, such as might be expected to occur in random text of the same length; hence they must be causal in their origin. The cause in this case is not difficult to find: repeated isolated letters and repeated sequences of letters (digraphs, trigraphs) in the plain text were actually enciphered by identical alphabets, resulting in producing repeated letters and sequences in the cipher text.

APPENDIX 3

A GRAPHICAL METHOD OF RECONSTRUCTING PRIMARY COMPONENTS BY APPLYING THE PRINCIPLES OF INDIRECT SYMMETRY OF POSITION[1]

1. **Fundamental theory.**—*a.* It has been shown that the interval between letters of a sequence obtained from a secondary alphabet is a constant function of the interval separating the letters in the original primary component. Consider the following sequence:

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z

Assume that this component is slid against itself and that the following groups of partial sequences are obtained from three secondary alphabets:

Group 1—S T I; U E; N A
Group 2—I N; E T; O A
Group 3—T N; Q S O

Figure 1.

Referring to the primary component, it will be seen that the letters of the partial sequences obtained from group 1 coincide in their interval with that in the primary component; the letters of the partial sequences obtained from group 2 represent a decimation interval of two in the primary component; and those obtained from group 3, a decimation interval of three.

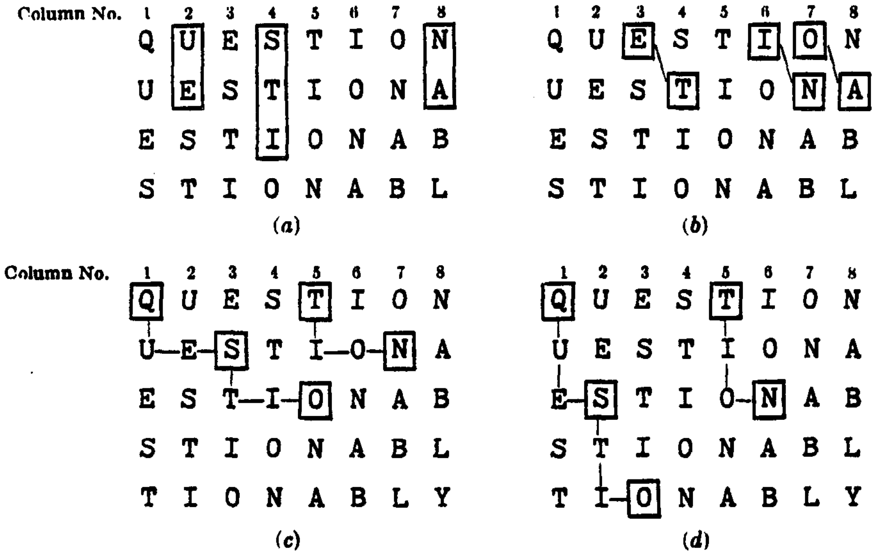
b. In the foregoing case, decimation was accomplished by taking intervals to the right along a horizontal component. Given Figure 2 below, let a portion of that square table or matrix be considered, as shown in Figure 3:

---

[1] The basic theory underlying this modified method of applying the principles was set forth in a brief paper (November 5, 1941) by 1st Lieut. Paul E. Neff, Sig. C. To his original notes, which I have slightly modified for purposes of clarification, I have also added the matter contained in Pars. 3*e* and *f*.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z
U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q
E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U
S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E
T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S
I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T
O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I
N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N
B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A
L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B
Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L
C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y
D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C
F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D
G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G
J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H
K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J
M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K
P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M
R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P
V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V
X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W
Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X

FIGURE 2.



(a) (b) (c) (d)

FIGURE 3.

c. Again referring to Figure 1, the partial sequences STI, UE, and NA can be obtained from Figure 3(a) by reading down columns 4, 2, and 8, respectively. This can be represented graphically by the symbol ↓1, which means that all partial sequences obtained from Figure 3(a) by proceeding downward in any column would be in the same group (i. e., secondary alphabet) and have the same decimation interval.

d. The partial sequences IN, ET, and OA can be represented graphically by $1\,|\underset{1}{\rightarrow}$, or simply ↘1, which indicates that all partial sequences obtained by taking letters one space down and one space to the right; or one space down a diagonal to the right would represent the same decimation interval.

e. The partial sequences TN and QSO can be represented by the symbol $1\,|\underset{2}{\rightarrow}$; but they can also be represented by $2\,|\underset{1}{\rightarrow}$ and, if the entire matrix of Figure 2 is considered, by other possible routes.

f. The decimation interval of a secondary sequence derived from a primary is the sum of the horizontal and vertical components of the route selected. Since the partial sequence TN can be represented by $1\,|\underset{2}{\rightarrow}$, the decimation interval of this sequence is equal to the vertical decimation interval of the basic square plus twice the horizontal decimation interval in that square. Any other route selected for the same sequence would give an equivalent of this.

g. It is seen, therefore, that the decimation interval of a component can be represented graphically in various ways other than along the horizontal, by use of diagrams such as in Figure 3, in which the successive juxtaposed components have the same relative displacement. In this case the successive horizontal lines had a one-letter displacement to the left.

h. Not being limited to one dimension, reconstruction of the primary component or an equivalent should be possible in one combined matrix by reversing the foregoing process and graphically integrating partial sequences from different secondary alphabets into a single diagram. Suppose the partial sequences in Figure 1 are given and it is desired to reconstruct the primary component.

Group	1	2	3
Partial sequences	S T I	I N	T N
	U E	E T	Q S O
	N A	O A	
	$\underset{1}{\rightarrow}$	↓1	↘1

FIGURE 4.

i. (1) Using cross-section paper one can arbitrarily select the STI sequence in group 1 and write this sequence horizontally, making the graphical notation $\underset{1}{\rightarrow}$ below group 1.

(2) Proceeding to group 2, the partial sequence IN contains one letter in common with the sequence STI already entered, but since NA forms a sequence in group 1 and OA forms a sequence in group 2, it is clear that two different decimations are involved and therefore it would be incorrect to integrate the STI and the IN into STIN. However, the letter N can arbitrarily be placed in any position *other* than along the horizontal line on which STI has been placed. It will be placed directly below the letter I and the group will be denoted graphically by ↓1, giving:

$$\begin{array}{ccc} S & T & I \\ . & . & N \end{array}$$

FIGURE 4 (a).

(3) The skeleton of the matrix or diagram is now fixed in two dimensions, and no further letters can be *arbitrarily* placed within it. However, additional sequences from groups 1 and 2 can be added, provided a common letter is available in the diagram; sequences from other groups can be added, provided one pair is already entered in the diagram which would fix the proper graphical decimation.

(4) Moving to group 3, there is the partial sequence TN and it is noted that this pair of letters is present in the diagram. The symbol ↘1 can therefore be placed under group 3.

(5) In group 3 the partial sequence QSO appears and the letter S is in the diagram. It therefore follows that the letters Q and O can be placed thus:

```
(1) Q . . .
(2) . S T I
(3) . . O N
```

FIGURE 4 (b).

(6) Similarly the letter E of the partial sequence ET in group 2 goes directly above the T:

```
(1) Q . E .
(2) . S T I
(3) . . O N
```

FIGURE 4 (c).

(7) The letter U of the sequence UE in group 1 goes before the E:

```
(1) Q U E .
(2) . S T I
(3) . . O N
```

FIGURE 4 (d).

(8) Likewise the letter A of NA in group 1 follows N:

```
(1) Q U E . .
(2) . S T I .
(3) . . O N A
```

FIGURE 4 (e).

(9) The sequence OA in group 2 remains to be entered. Since both these letters are already in the diagram, the letter A can be placed under the existing O or the letter O can be placed above the existing A. Either alternative would be correct. Selecting the latter alternative yields the following:

```
(1) Q U E . .
(2) . S T I O
(3) . . O N A
```

FIGURE 4 (f).

j. All the original information has now been entered in the diagram seen in Figure 4 (f) and the letter O appears twice therein. This letter O may be termed the "tie-in" letter since it indicates the horizontal interval between the juxtaposed reconstructed sequences of the basic matrix. The absence of a tie-in letter in the diagram would indicate that insufficient data are present for the reconstruction of a complete sequence.

k. (1) By sliding the last row of Figure 4 (f) two intervals to the right the two O's can be superimposed, giving:

```
(1) Q U E . .
(2) . S T I O . .
(3) . . . . O N A
```

FIGURE 4 (g).

(2) Since each horizontal sequence must be shifted two intervals to the right of its initial position in relation to the line above, row (1) must be moved two intervals to the left of its original position. Thus:

(1) Q U E . . . . . .
(2) . . . S T I O . .
(3) . . . . . . O N A

FIGURE 4 (h).

(3) Since the three rows involve the same decimation, and since the O of ONA coincides with the O of STIO, the ONA sequence may be raised up one row and united with the STIO sequence. If this is legitimate then the new row (2) may likewise be raised up one row. This yields the united sequence QUESTIONA. . . . This last step may be more clearly understood by studying the following partially reconstructed matrix:

(1) Q U E S T I O N . .
(2) E S T I O N A B . .
(3) T I O N A B L Y . .
(4) O N A B L Y C D . .

FIGURE 4 (l).

2. **Application of principles.**—*a*. For the specific application of the principles underlying this method reference is made to the problem described in Section VIII of the text. It is desired to reconstruct the original primary component, or an equivalent, from the values entered in the reconstruction skeleton shown in Figure 33, page 68. Since a mixed sequence is sliding against itself, all the partial sequences (pairs or greater) which can be established by studying the reconstruction skeleton are listed as shown in Figure 5(a). The single pairs in Ø–7 and Ø–8 are crossed out since they offer no data for reconstruction. This yields the following groups of partial sequences:

Ø- 1	2	3	4	5	6	7	8	9	10
BW	EK	EX	AE	ED	EJ	~~UO~~	~~GC~~	IHJ	HE
EGZ	HZ	TU	HG	HCR	GN			TS	IV
GZ	NS		IO	NP	HOF			WA	NQ
TK	UF		TP						

FIGURE 5 (a).

b. (1) The sequences HOF and EJ in group 6 and HE in group 10 are noted. The HOF will be placed horizontally and the notation $\underset{1}{\rightarrow}$ made under group 6. The letter E of the pair HE of group 10 will be placed under the H, and the notation ↓1 added under group 10. Thus:

Ø- 1	2	3	4	5	6	7	8	9	10
BW	EK	EX	AE	ED	EJ	~~UO~~	~~GC~~	IHJ	HE
EGZ	HZ	TU	HG	HCR	GN			TS	IV
GZ	NS		IO	NP	HOF			WA	NQ
TK	UF		TP						
					$\underset{1}{\rightarrow}$				↓1

FIGURE 5 (b).

Since the sequence EJ belongs to the same displacement interval as HOF, the letter J can be inserted after the letter E, giving:

H O F
E J .

FIGURE 6 (a).

No more pairs can be immediately added from groups 6 or 10. Those pairs already entered are crossed out in their respective groups and an inspection is made for additional data in another group.

(2) The sequence IHJ is noted in group 9. The letters H and J are already entered in the diagram. One can therefore place the letter I, and the notation $\searrow 1$ is placed under group 9. The addition of the letter I now permits the insertion of the letter V of the sequence IV in group 10, giving:

```
I . . .
V H O F
. E J .
```

FIGURE 6 (b).

(3) In group 4 there is the sequence IO which is obtainable in the diagram by the route $1|\underset{2}{\rightarrow}$. This notation is made beneath group 4; the letter A of the sequence AE and the letter G of the sequence HG can now also be entered. The addition of the letter A permits the placement of the letter W of the pair WA of group 9; likewise the addition of the letter G permits the insertion of the letter N of the sequence GN of group 6; finally, the placement of the letter N permits the placement of the Q of group 9. One now has:

```
W . I . . .
. A V H O F .
. . . E J G N
. . . . . . Q
```

FIGURE 6 (c).

(4) Referring to group 1, the sequence EGZ is noted, of which EG appears in the diagram at $\underset{2}{\rightarrow}$. The letter Z can therefore be placed and the letter B of the sequence BW can be inserted two intervals to the left of the letter W, giving:

```
B . W . I . . . . .
. . . A V H O F . .
. . . . . E J G N Z
. . . . . . . . Q .
```

FIGURE 6 (d).

(5) Noting the sequence HZ of group 2 as being graphically represented in the diagram by $1|\underset{4}{\rightarrow}$, the letters K, S and U of the sequences EK, NS and UF may be placed. Thus:

```
B . W U I . . . . . . . .
. . . A V H O F . . . . .
. . . . . E J G N Z . . .
. . . . . . . . Q K . . S
```

FIGURE 6 (e).

(6) The letter T of the sequence TK of group 1 can now be placed, which permits the addition of the letter P of the sequence TP of group 4. A study of the diagram shows the pair TU of group 3 at interval $3\underset{4}{\uparrow}$, which allows the placing of the letter X of the pair EX of the same group. One then has:

```
  . X . . . . . . . . . . .
  B . W U I . . . . . . . .
  . . . A V H O F . . . . .
  . . . . . E J G N Z . . .
  . . . . . . . T Q K . . S
  . . . . . . . . . P . . .
```

FIGURE 6 (f).

(7) The diagram now shows the pair NP of group 5 at $2\underset{1}{\mapsto}$. The letter D of the sequence ED and the letters C and R of HCR can therefore be inserted. Thus:

```
  . X . . . . . . . . . . .
  B . W U I . . . . . . . .
  . . . A V H O F . . . . .
  . . . . . E J G N Z . . .
  . . . . . . C T Q K . . S
  . . . . . . D . . P . . .
  . . . . . . . R . . . . .
```

FIGURE 6 (g).

(8) Pair TS of group 9 remains. It has already been noted that the notation $\searrow 1$ has been applied to group 9. Hence the letter S can also be placed one interval to the right and below the T, as shown in Figure 6 (h), in which all the available data are now entered.

```
(1) . X . . . . . . . . . . .
(2) B . W U I . . . . . . . .
(3) . . . A V H O F . . . . .
(4) . . . . . E J G N Z . . .
(5) . . . . . . C T Q K . . S
(6) . . . . . . D . S P . . .
(7) . . . . . . . R . . . . .
```

FIGURE 6 (h).

c. (1) The letter S appears in rows (5) and (6) at a displacement interval of four. This letter then serves as the "tie-in" letter. Marking off 26 squares on cross-section paper the D.SP of row (6) is written, and row (5) is moved four intervals to the left, at which position the letter S is properly superimposed as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Row (5)	C	T	Q	K	.	.	S	.																		
Row (6)	.	.	.	.	D	.	S	P																		

(2) Likewise row (4) is moved four intervals to the left of its original relative position to row (5) and dropped into position. Row (3) is moved the same distance in relation to row (4), etc. These steps may be illustrated as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Row (4)	.	.	.	.	E	J	G	N	Z	C	T	Q	K	D	.	S	P	.	.	.	.	.	.	.	.	.
Row (3)	H	O	F	.	E	J	G	N	Z	C	T	Q	K	D	.	S	P	.	.	.	.	.	.	.	A	V
Row (2)	H	O	F	.	E	J	G	N	Z	C	T	Q	K	D	.	S	P	B	.	W	U	T	.	.	A	V

(3) The placing of the letter X of row (1) and the letter R of row (7) gives the final sequence:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	O	F	.	E	J	G	N	Z	C	T	Q	K	D	X	S	P	B	R	W	U	I	.	.	A	V

(4) It will be noted that the foregoing component is identical with that obtained in subparagraph *m*(3), page 74 of the text.

3. **Remarks.**—*a.* In the example given above only one tie-in letter was available and it was located in adjacent rows. Although only one is necessary, in most cases several tie-in letters are present after all pairs of letters have been entered in the diagram; then the superimposed sequences can be easily connected by their common letters. If the tie-in letter had appeared in adjacent columns instead of adjacent rows as in the foregoing example, the columns would have been shifted vertically and the sequence taken from the diagram in that manner.

b. When only a few pairs of letter forming partial sequences are available, frequently only one tie-in letter may be encountered. If it does not occur in adjacent rows or columns the component can still be written with additional considerations. For example, adjacent diagonals might be used. However, the student will experience no difficulty after the application of this method to a few problems.

c. Since all the data are entered in one diagram, the graphical method of reconstruction quickly discloses erroneous assumptions and enables one to ascertain in a short time whether sufficient data are present for the reconstruction of the component. Even if this is not the case, the diagram automatically offers new values which may be substituted in the cryptogram. One may then assume additional values which can be entered in the diagram or which will serve to corroborate sequences already entered.

d. The placing of the first two sequences of different displacement intervals in the diagram determines the type of sequences that will be established. If the original sequence entered horizontally in the diagram is an odd decimation of the primarv component, a 26-letter sequence can be obtained horizontally. If this original sequence is initially tied in vertically with another sequence of an odd decimation interval, a 26-letter sequence can also be obtained vertically from the diagram.

e. (1) In certain instances, however, it will happen that the available partial sequences have all resulted from even decimations of the basic sequence and that no tie-in letters are present to permit the integration of all the data into a single diagram. In such cases the reconstruction of the basic sequence may take place by taking data from two or more different diagrams, and then, using the relative positions of the letters with respect to each other in these diagrams, the basic sequence may be established. This method can best be demonstrated by means of an example, and the following one is based upon the QUEST . . . sequence of paragraph 1*a*. Suppose the reconstruction diagram from the derivation of a few plain text-cipher relationships yields the following partial sequences:

Group	1	2	3
Sequences	Q H O	Q X V	Q T A
	F T	O T	X E
	C E	P K	F K
	J N	F C	U I B
	W D S	U Z W	Z S
		N I	Y G M
		G D	

FIGURE 7.

(2) The partial sequences in the three groups can be combined to form two diagrams. This may be accomplished by considering the sequences of group 1 as parts of a horizontal component and those of group 2 as parts of a vertical component of a cipher square based upon the original or an equivalent primary sequence. When all the letters of these two groups have been entered into the two resultant diagrams in Figure 8 (a) and (b), it will be observed that the positions occupied in these two diagrams by the letters of group 3 represent the interval $1\underset{2}{\lfloor\!\!\rightarrow}$. Thus:

(a)	(b)
Q H O . .	Y U J N . .
X F T P .	. Z G I . .
V C E K A	. W D S M B

FIGURE 8.

(3) It will be noted that there are 12 letters in each of the two diagrams and that all the letters appearing in the original partial sequences have been included in these two groups. It appears, first, that two 13-letter sequences are involved and second, that the partial sequences in all three groups represent even decimations of the basic component. The problem now remains to reconstruct the original or an equivalent primary cipher square to which these diagrams belong, or to find the original or an equivalent component of which the partial sequences in groups 1, 2, and 3 are derivatives.

(4) Since the two diagrams are linked by the partial sequences of group 3 (because the interval $1\underset{2}{\lfloor\!\!\rightarrow}$ is common to both of them), it follows that any two letters in one of the diagrams will be separated from each other in the basic sequence by the same interval as any two letters occupying the same relative positions in the other diagram. Another way of saying the same thing is, that while the intervals between V and C, C and E, E and K, and K and A, in the basic component (or an equivalent thereof) are unknown, whatever they are they are identical and the same as that between W and D, D and S, S and M, M and B (from WDSMB), or between Y and U, U and J, J and N (from YUJN), and so on. Likewise, Q and K (interval $2\underset{3}{\lfloor\!\!\rightarrow}$) are separated by the same interval as Y and S, or U and M, and so on.

(5) Making the easiest assumption first, suppose the basic sequence is a keyword-mixed sequence, and that the letter Z is the final letter thereof. If it is preceded by Y, then, because of the relative positions occupied by Y and Z in Figure 8 (b), the following would also be sequent in the basic sequence: QF, HT, OP, XC, FE, TK, PA; UG, JI, ZD, GS, and IM. Since the majority of these are hardly likely to occur in a keyword-mixed sequence, the assumption that Y precedes Z is discarded. Suppose X precedes Z (implying that Y is in the keyword). But X and Z are not in the same diagram, so no test can be made. Suppose the sequence is W . Z. Then the following sequences would be valid:

W . Z . U	V . X . Q
D . G . J	C . F . H
S . I . N	E . T . O
	K . P

These look very likely. In fact, noting the D.G.J and the C.F.H sequences it seems logical to integrate or "dovetail" them thus: CDFGHJ. This then suggests that W.Z.U and V.X.Q may be integrated into VWXZQU; S.I.N and E.T.O may be integrated into ESTION. From this point on the matter of extending the partial sequences into the basic one is simple and rather obvious.

f. (1) Suppose, however, that the basic sequence is not a keyword-mixed sequence, so that clues of the nature of those employed in the preceding subparagraph are no longer available. Then what?

(2) Referring back to subparagraph *d* (3), it has already been noted that the two diagrams, each containing 12 letters, represent half-sequences (of 13 letters) derived from an even decimation of the original component. (The decimation must be the same in both cases because the interval $1\underset{2}{\lfloor_{\rightarrow}}$ is common to them.) Suppose an attempt is made to integrate the QHO, XFTP, and VCEKA sequences of Figure 8 (a) into a 13-letter cycle or half-sequence. The three partial sequences in this diagram may be united into a 13-letter cycle in a number of ways but the correct integration will be that which will satisfy all the conditions set up by the partial sequences in groups 1, 2, and 3. After a bit of experimentation it is found that the only one which will satisfy all conditions is this:

1	2	3	4	5	6	7	8	9	10	11	12	13
Q	H	O	V	C	E	K	A	X	F	T	P	.

Note, for example, that the conditions represented by QXV in group 2 are satisfied in that the intervals between these letters are the same in the 13-letter cycle; the same is true as regards the intervals between O and T, P and K, and so on. Likewise, the conjugate sequence from Figure 8 (b) is established as

1	2	3	4	5	6	7	8	9	10	11	12	13
Y	U	J	N	W	D	S	M	B	Z	G	I	.

Thus there have been established the two half-sequences involved. The problem now remains to integrate them into a single sequence which is either the primary or an equivalent basic component.

(3) Each of these sequences may, of course, be expanded to form a 26-element sequence the elements of which will satisfy the interval relationships among the letters in each 13-letter sequence. Thus:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
(1)	Q	.	O	.	C	.	K	.	X	.	T	.	.	.	H	.	V	.	E	.	A	.	F	.	P	.
(2)	Y	.	J	.	W	.	S	.	B	.	G	.	.	.	U	.	N	.	D	.	M	.	Z	.	I	.

FIGURE 9.

There remains the problem of integrating these two sequences into a single sequence.

(4) Suppose a start is made thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Q	Y	O	J	C	W	K	S	X	B	T	G	.	.	H	U	V	N	E	D	A	M	F	Z	P	I

FIGURE 10.

All the interval relationships of groups 1, 2, and 3 of Figure 7 are satisfied by this sequence. If the sequence is written on a pair of sliding strips, any even-interval displacement of one of the strips will produce plain text—cipher relationships fully satisfied by the requirements of the sequences in Figure 7 or Figure 8. Thus:

```
(1) Q Y O J C W K S X B T G . . H U V N E D A M F Z P I
    H U V N E D A M F Z P I Q Y O J C W K S X B T G . .

(2) Q Y O J C W K S X B T B . . H U V N E D A M F Z P I
    X B T G . . H U V N E D A M F Z P I Q Y O J C W K S

(3) Q Y O J C W K S X B T G . . H U V N E D A M F Z P I
    T G . . H U V N E D A M F Z P I Q Y O J C W K S X B
```

FIGURE 11.

The foregoing three juxtapositions will satisfy all the requirements of the sequences indicated in groups 1, 2, and 3 of Figure 7, as well as those indicated in Figures 8 (a) and (b). Without further restrictions or additional data, therefore, it is impossible to tell whether the reconstructed single sequence is correct or not. In fact, there are 13 possible integrations of the two expanded 13-letter sequences which will yield equivalent results, since there are 13 positions in which the "dovetailing" of the second sequence may be commenced with respect to the first sequence. Only one of these, however, will be correct in that it will yield a single sequence which, when slid against itself at all juxtapositions (both odd and even displacements) will invariably yield the full quota of plain text—cipher relationships that the original basic or an equivalent primary component yields when slid against itself. (An incorrect integration will often yield a series of equivalents of which only a few are wrong.)

(5) The correct integration will, however, be disclosed quickly enough when the cipher text is consulted and one or two additional values are derived. Thus, suppose an additional word is deciphered and it yields a pair of values in a new secondary alphabet, for example, $A_p=D_c$ and $U_p=O_c$. The single sequence reconstructed as shown in Figure 10 will not yield this pair of values, as seen in the following juxtaposition of the sliding strips:

```
Q Y O J C W K S X B T G . . H U V N E D A M F Z P I
I Q Y O J C W K S X B T G . . H U V N E D A M F Z P
```

FIGURE 12.

Here $A_p=D_c$ but $U_p=H_c$, not O_c. However, if the "dovetailing" is commenced with the letter S, of Figure 9 and the resultant 26-letter sequence is juxtaposed against itself as shown in Figure 13, it will be found that the sequence will now satisfy all the requirements.

```
Q S O B C G K . X U T N . D H M V Z E I A Y F J P W
I A Y F J P W Q S O B C G K . X U T N . D H M V Z E
```

FIGURE 13.

The sequence is, of course, a decimation of the QUESTIONABLY . . . sequence, at the third interval.

INDEX

Page
Accidental repetitions 12
Alphabets:
  Classification of 4
  Derived 4
  Interrelated 24
  Mixed 24
  Secondary 4
Analytical key 95
Aperiodic systems 2
Assumptions for values, check 76
Average 109
Bazeries 23
Beaufort 9, 19
Causal repetitions 12
Cipher disks 5
Classification of alphabets 4
Coincidence:
  Digraphic 115–116
  Monographic 113, 114
  Tetragraphic 117
  Trigraphic 117–118
Comparisons, number of 113, 116
Completing the plain component sequence 19, 79, 82, 88
Component monoalphabets 15
Constant intervals 85
Conversion:
  Into monoalphabetic terms 92, 94
  Into plain-component equivalents 81, 83
Cryptograms:
  In different keys, containing identical plain text 85
  With plain text 84
Cyclic phenomena 2
Data pertaining to:
  Digraphs 115–116
  Trigraphs 117–118
Decimation Footnote 1 53
Delastelle 9
Derived alphabets 4
Digraphic coincidence 115–116
  Probability for 115
Digraphs, data pertaining to 115
Direct symmetry 9, 26
  Application of principles 32
  Of position 84
Distribution of different categories of letters in respect to frequency of occurrence 109
Double-key system Footnote 2 3

Page
Equations, enciphering 5, 6
Equivalent primary components 53
Expected number of occurrences 109
Factoring 15
  Intervals 86
Givierge Footnote 1 23
Gronsfeld 21
Identical messages enciphered by keywords of different lengths 89*ff*
Identical superimpositions 86
Index letter 6
Indirect symmetry 8–9
  Of position 52–77, 119–129
Interrelated alphabets 24
Latent symmetry 52
Law of small probabilities 109
Matching 47
  Distributions 94
Mean number 109
  Coincidences, of 114
  Digraphs, of 115
Mixed alphabets 24
Monoalphabetic terms, conversion into 46
Monographic coincidence 113*ff*
Multiple alphabet system 3
Number of comparisons 113–116
Parameter Footnote 2 113
Partial chains of equivalents 85
Period, determination of 10, 15
Periodic systems 2
  Primary classification 3
Phenomena of repetition 108
Poe, Edgar Allan Footnote 1 2
Poisson's exponential expansion 109
Polyalphabetic substitution:
  Distinguished from monoalphabetic 1
  Primary classification 2
  Sequence of study 3
Primary components 4, 5
  Equivalent 53
  Reconstruction of 27, 52
Principles of indirect symmetry of position:
  Application of principles 69*ff*
  Application to specific example 60*ff*
  Fundamental theory 68, 69
Probability:
  Definition of *apriori* 108
  Of digraphic coincidence 115–116

Page

Probability—Continued.
  Of monographic coincidence 113*ff*
  Of tetragraphic coincidence 117
  Of trigraphic coincidence 117
Probable-word method 21, 43
Progressive alphabet system 3
Random text 108
Reconstruction of equivalent primary components 4, 5, 27, 52, 53
Reconstruction skeletons .... Footnote 1.. 26, 56
Relative frequencies 40
Repeating-key ciphers:
  Primary components are different mixed sequences 80
Repeating-key system 3
  Analysis of 10
  Solution of subsequent messages enciphered by the same primary components 78*ff*, 80*ff*

Page

Repetitions:
  Accidental 12, 116
  Causal 12
  Phenomena of 108
Secondary alphabets 4
Sequence reconstruction skeleton 26
Square tables 5
Symmetry of position 8, 9, 52, 119–129
Tetragraphic coincidence 117
Theory of factoring 15, 86
Trigraphic coincidence 117
Trigraphs, data pertaining to 117
Types of cipher squares 96*ff*
Vigenère 9, 19

○

PROBLEMS

VIGENERE Ciphers. Text is military and key-lengths are between 4 and 11.

1. Q N V K J   P N W D O   C Y N T V   Y U T P F   L S V Q O   R V Y D E
Q L Z T Y   C W A F W   F E E B L   Y F L Z D   Z A L T V   Q S T Y Y
E A A R R   F V L D I   I H A P V   P L T V R   Z C R L M   E R A B E
T R P D N   W M N P R   T M R V F   O G U T L   Y N Q E G   W Z U N G
H I J C O   I S P A R   V B P F V   Q U T V G   Y F B U Y   A K G Z N
V V D E T   C T V R G   (160)

2. K O F W A   D A Q L W   E G R N T   Z C H Z T   Z F W P D   R R D K V
L T H Z S   X I P M B   K S W W D   S E S Z S   G A U M R   K O P W J
V I Q B V   V N H F H   K W H V H   P F R C F   Y O X Z G   J T R X M
F U U Z S   X I P M B   K W L T Z   S E D B H   R C K M R   K O W P S
J E F W B   U D L D W   K I R V   (134)

3. H I B K M   I M R T M   W L W E B   T J Z V J   C P M Y J   C I G U Z
C N D B Z   U V P W N   N C D V B   V W N Y C   I O V T B   H Z V V D
B J M E F   A K U X J   M A M M G   G Z G Q C   P T R X T   L A H J Z
K U N F M   M D I L Z   W Z I T D   K I O N Y   M A H U L   T Q G Q O
I A K U X   F Q G I M   U M M A M   R V T U G   D V L S J   V
(146)

4. N I P Z H   G V I I N   M W I I O   S N G L P   N E P L T   Y V Z R I
I G M S I   P C E B N   L K D S U   K Z W K K   T T A P W   N E P L V
V Q G H I   T X F S L   N T E K Z   S E J I L   K L U L E   W I M P S
E Z E T X   K J Z R I   A I Y M C   O X O I H   H M E W E   P O I E P
L E R O E   M A E R K   P Z V Z   (134)

5. G L A P S   Z C U B K   M A K A T   W U O P K   X M S X D   C S E O O
G Y G S X   D B W M S   U Q H L O   I E L U I   O R T T A   R H X W H
Y O V E X   L G C O C   J V X V P   W G I M O   G C E C V   W I Q V R
Y C H Y O   V E X L G   C O C J V   X V P B Q   N Q A A V   R X K H X
B Q M R G   G Z Y P E   X Y E I W   R H R M J   M W Y F I   L Y F R D
G B T L X   H A Q H I   Y A M K Z   N X P Y Z   Z Y P B H   W B Y N S
B Z D   (183)

6. WPOQL WXJTF YOZMK YYPLI LRXZY SGOOI
EOVFY NAIVD KZPOW IGTFW GJAIP WNRCJ
JIXCF TLEUK WETAR DLXVL NYGJN AMWQN
MITIR NJZLO PBWPU ECIII VLSOU ESPPL
XEDKY WMIBP CGZ (133)

7. LEARJ PRAIK ICTDH ESAGI LFBPG OYNOC
HTHGK PWAOV ALSET EWYXT BRZTC TXIFZ
VAMZC YPNWK VPXNM VFCDG TKWZC RXGCC
IWBCZ MTSXE HJDUK YSLSQ NRFET RLKVT
HAYKS CCOHE GEDP (134)

8. JIWBX YUCLA GCTRH SGYQX PBCCM YGXCY
EOTPG NMOVD SYNKD NUOYN GXNYK CFHTO
LXCLW YUDZJ RVPUU OCYRO LNVRP FCCEM
GXEYP MP (97)

9. RYWOW LTNNU LROOX YHXAN EIMKM YYMSZ
EQNCU ACITN SBNMZ BWXXN KLFTL GKMKT
GYTOE UELYC PMEYT XSMEX LIMKJ ORSKM
OJZBW FEMYU YXCIT NSBNM MUJUL YYUFW
IMAHJ XAXGV DX (132)

10. HPVEJ ZHTXZ HLPPY YIGKV LUXPH SMFEU
XBJVE EOMLV GCMXO ZFPIV DQXRW TKHVU
TLRIS ZBAVV VMBQA ELTXZ HLHGE KMWME
WCEEN KUXRK KCVYH RBBSE YMJVB ZDXTF
BLIWE OAMSG KCFYA JBTRF YDXGA IJXEJ
LGVRA UBAMJ ACPHM LIKXV KQXRY FVGIT
MGDRS ZBAXY XAWEJ XML (198)

11. ULZTY HASWD CGTMR GPRDQ ZCASH MVHDM
DHFUE LTVIX OKTQY XNLRL FHDIH NYTMP
JYIXG IGMKD AVTKL TDXEY IBTWH CFCQS
GILGL HKUER PTPLK NESRL FHDXW UKQCM
SAV (123)

```
12. OHTVT CHHBW OHTHE KGHEK KSYHS MTVGH
    EHSDE LTUER GJZLJ QZQZO NWCWE KUICF
    UAMDZ DLWXC WVYHZ SZVPU KODRT TZUZV
    TVGQE LUBTV ABVQU KYVZC A     (111)
```

TRUE BEAUFORT Ciphers. Text is military and key-lengths are between 3 and 10.

```
13. VWREC TCSOL AWOQC GWKBU KTEUP TAYVK
    IDNEL ZAVAP JSEIJ YJOAF HSLEO ZWOPO
    XHZEQ IIOIW ISOUP BCRIL TSGXR OJCIZ
    MZVEK TMR   (98)
```

```
14. HXCJO BDROW UZPJV KDRHS EMSZQ MYMWW
    ESVVS CWUEE SSZLK ASGJG LPRYS YKTHZ
    WQGXE WMPZQ AZWQA KVFIA SWUQF HZOIN
    AUAXS WANDL WWKUN ABJHD QKACH ARLXR
    FAAAE FORRP HFPRY TDMTF JQWEU TLJLV
    VHYTK EEHNS WDNDM KMRRQ WSYYT AVKEE
    XPWMP NLMJI SGJGV PQE   (198)
```

```
15. WWMJW PHIUM ALEIW ZBIAA AODWU ENQLA
    OXCMK KWAAJ BDOXV VRDSE AWUJT RJAHI
    CRVKA KIRWP SEDUY HGDPI CVVDL ECJMM
    AEVWO LLBNI KEIWJ OXOEU AZXEE BZMVL
    YQUPZ ASUIB ACFQJ HGLLY JZZYI ONATO
    EEK   (153)
```

```
16. EPNWV VAPIZ YPAZG MFCAT ZAGLG YTLZC
    TGOLN KATNG BSKPZ QMPMO NQRKA ZGJCZ
    EDZQK CJCPZ GKCGO KGVLY CMONQ R
    (86)
```

```
17. FEXZA ZDIWA WGLEW VMXHT ZCLQS TWFJJ
    TEKTM ZIUAT MCIQV LDGSQ UFPVS IWPLR
    MAANW EUUBL RLKUW AEAFU TGBZS NJCGK
    IOWBL DHMJT TZIGN GKZSI FCMXH TZCL
```

18. QDCCA LGFGI XXYNN QBTOF EKRNX DTVDN
BEMNU NTPEJ WCPAI LJENN ALVJI NTAAW
EELEU WKTGP MTRYM NFBYM PBPRJ AKFVL
TXGES RRFEO QCLEU QDCRX JXGPW CJHFA
RUEDI LEFAM EUKXN CELDN IQLZQ AZARN
YJOJF AFLUH IJAVM (165)

19. ADMDT BDJNG OMTHR KPDVT VNFLG VZPMV
NZGHN BGGZM JYYEL ATKAU RNTWA HYZGZ
UYXRY KZJYB YRLMX GVGAV LJDLT VIDGV
VNYMD UGAHI LZYWU KPDVE HNZLC AHNJY
VDEYB WNDLQ VNLIP OPASR YZYMY GONTE
QYRPA ICEWV CGSPA KOKAU BNUCA XIJNG
JDHLT MZKSA (190)

20. AUQFK OFXUO VDMCF KSKIY ORXYV RANQB
VLBOZ ZLVBH YATQW IYLDZ ZEAAM KXBBW
YGGPK VXJID ETVGO JEKDZ CGVQM ITKCF
KSKIL YQKHH USHBW YBKWH OULEJ OSKTS
IHUHN IXVQ (129)

21. HQJCD NOMZZ MCNHL NAAYM CZMQK AAODJ
ONZXD VYGBD DWDYV YGYXC JGUAO JFEQL
DIBXH FPUUJ GAZAA YMCYN NYBFF QVYGY
(90)

22. KPYVG UMOVN UTLZU UMLMZ DHENJ HEOEG
ETLLZ KUTZV EKHXN RMGKX EHJJX TORSM
CYEEV NZOXB BUKVH XDKDI ZGYOO FGVID
AZITL WLPXA OJSZG OFYNH HHUIT KEHWK
XVYJX XKFCZ NHRSP NNPDE FUBUT HZOVN
GLLBS GZTHH BOCYB WYZZU GVIDA Z
(176)

23. DEAZG RATBA QWSQC QMKDX WXVNB APZDQ
EKOTO LBCLP AYGPJ AKKTC BQBTC GUMQR

NKROF MKOOP SQNTE LBCCL NQNXP GMSQQ
IAVWS ASMTA AWSUQ EKMUD KUND (114)

24. ZIGYA WNTPO VGGEA CBWXH YOWNL TAREQ
FSFEP YXKZM BCVET MFVHK URIDW SGOPU
XRJOO JPLBU LWKJC XYQFN KIVTY QJLOJ
PIIVU ZQIRX UQUSD AABBI KLKEE OYOVZ
EFPBO GZTQS RZEVH NKKEY HWZLO KEZTQ
USDDB EUTWC HOWEZ IWIKJ DYVWN NKIVE
UGCGA EAHEE KOOUR IT (197)

VARIANT BEAUFORT Ciphers. Text is military and key-lengths are between 3 and 9.

25. MTNZA BECPA LFAWJ ZBXAO BNRTE MKCPW
MFQMA GQVGS BXUJA ZUWBD XUAMT IQLBA
WAONA GENQJ MTNVA QFOWN MKNQC AFQWQ
KEBBK IUOMJ XYHIN MUUTA KKOQN XEDLZ
XZUGE GOAMW LQBIZ OUBMP AUBPA TPZCW
KFNZO BYVMZ BMCMH R (166)

26. OSYDD JCULA QKHRS CUODA BFGTM SABUL
KUBWC QROMS PQQCU QLUQM QVZKU NZMBS
GWVYS UUSEX ZDVDS SXICL WROCA NPRSD
NDCHD DWWKA OMMCN EGUFG ANERP YZPNE
ULIHH UMBCU RAHAS PYLAM UYAQK U
(146)

27. AFTQF TYQUU YWFWG MAGOJ PRLLM KHTGI
PGLZL LAEXX WHQSH FGUTA AXGCQ KTHWC
LWGAR TDGHF GEFZT BAAAL TVWRF WSMFS
TSHNW TZUPB LJDWS (105)

28. UQPOC ODKAQ CQPMI FLSDR PDTPC DKBDA
GZCED KRTTP MGIEZ MCHMR IECNA RCDXS
CDHQM BDRPY PFLRD OAQHX PQSRE UCLLM

A F L Q G   R L Z X K   E R G L K   T X I J F   L L R T G   B C L T O
M L T O M   F D O M M   S G S Z R   Q G A C P   C H T K D   U K B Q U
M C G H T   (155)

29. T N G L M   O M E E K   P J C N M   U L Q I C   E W Z V E   B E K P Z
B M I X D   A O R B Y   Z K G N P   L Q K U D   F Q W A J   H S I Z A
M A B F A   K M B V B   F A X A K   H W D T X   H A R T A   B O Q F A
P C H U V   R M W Z X   A K J Z L   B J N F W   M N W T M   T A X L X
I A Z U W   C T Q K H   Q A U B Z   X A Q M B   V B F A B   G V Y K A
F X X B P   O H B O Q   L N L B Y   C V N M C   P T N T N   G V Z N D
R W J L M   K T X Q O   B   (191)

30. H K Q L K   A C K L P   X R P J J   Y X J E K   U W X C P   T L P G A
A U A E U   N Z S C S   X N E K N   Y I D N U   A A H P S   G C E W A
V K H H D   N O A B P   S E Y E X   F G Y C E   Y E D O J   Y C P B O
T D C P T   J S G B M   T A Y I X   X E N Z S   R D T Z D   Z X A R A
L Y N F H   E Y M C S   P P D E B   E T I T S   V P S E L   Z N W N
(149)

31. R A Q Z S   T W W B U   B R M L J   N P M Z E   N P N C C   Q B E N J
K A N E G   E U W S N   F M J P G   K R R Q Q   R E N L B   R P E H C
K N E V C   Z H E N H   P C F Q W   C I Y U Y   R A G J Y   A A V Q U
L E K L J   Q Y E P P   K U L A M   N O M R F   Y X K B F   Z S P O N
B E H Z V   A D N W T   O G K J A   J Q Q R F   Y P J P R   T O E Y F
O C C Y G   E U W   (158)

32. V Z M P Q   D G D K A   R H L M J   C D F U V   M W X A G   N Q T Z D
T I J K A   B Q U C X   O M F G D   R D C D Z   Q G D R D   O Z D A O
N U G H O   A Y V Q R   D L N F A   W C R D E   U A Z L R   T G P R N
Q Y Q Y L   X G X N L   T S T S A   A D A G E   S G J B E   A N P X U
O T Q Y D   I H P I K   O O Z N V   T T A G D   H L A Q Z   M P C H S
T S   (152)

33. S C X W H   W B Z P C   E X E E X   Z Y C L G   I K T D U   U B X H Z
X R K P P   M N B X P   P M R A Q   Z R D A E   N X W P V   A Q P B Q
N R A U O   B T F A Q   M A L U X   M N C K Q   Z S V U J   E M T A L
N Y X C K   A B A H B   E C L A B   R A Q O S   F Y W P R   F K R A D

```
    ACLFK  NIXAY  LQWBO  UOCPT  ALMTA  WTDNG
    OQ     (152)
```

```
34. IYETN  RZRPP  IZYDB  AACXO  OYPGK  FHLGE
    NOTRW  TPDID  AEEWA  EYPBU  SQTGO  TOTKE
    STZCE  SQLRE  NRJDQ  RQZJN  RPRXI  EYEHP
    OANPJ  YZFRK  NQTGI  TSTHK  ROPGK  FMLIP
    LPTSA  NETUE  CLEXK  NBFTN  Y      (141)
```

```
35. PPPCS  LCWYL  MPTGJ  DKDZM  HOTAT  EIPVE
    LLAWQ  PLCAY  LKAJA  GWJEL  DPWJP  EUQJK
    XKATR  EYYDT  DHEZN  PWBTN  AKWNZ  RACGO
    EGLLV  RTKAH  ZWEEA  LKQCW  OJGQL  JAESG
    TFCBM  ACQ    (128)
```

```
36. SAOEU  TMWZH  ZSWMU  QAWKR  AKZAT  JNLKE
    JWWOD  NQQZJ  WYPNN  NRNQB  AXNEM  RAHUH
    FRKMP  AUAUA  CEMPU  SIAQC  DDBDU  ALZCH
    NOMJQ  VZXBS  TQUVL  YHOEU  RJPWM  AWBTG
    JMYQP  ELQZJ  AAAYN  ZPUEQ  UFNNB  QBJAP
    ZMEBM  FUBOM  CESUE  DWAZX  SNZXY  VLECK
    OOMDG  VGYNN  HUTMH  ZHEMR  AHUHF  RKM
    (208)
```

The following problems concern Repeating Key Systems with standard and reversed standard alphabets. Read the messages and recover the keywords.

```
37. UCGZD  FRCAJ  GTZVF  WFZVU  KHZVH  DRUEW
    AZFRU  QGNBS  QCOJL  DZGBY  WHICR  KWNVR
    FGIAV  GINUP  GIHGD  ABNBP  GFLBZ  UCGZD
    UCGCO  WHCAJ  ECPRE  QHYAS  EGNBS  JSXSR
    JQYFH  KHCZD  LSXNW  SFYVQ  XCLPH  VRCIL
    KWIAD  JSGBY  ABABQ  YSNGB  KPOEJ  VOMUK
    SBIIH  JFING  KHICW  ZWMQL  NWMVR  FKCYO
    UCHGL  FIYGR  YIUEG  LVYEL  YVNSO  SBEBI
```

GILPR JDMFW GDUZP MBCGL GBQVO DQIAW
ABORW GPYSX JBCFK WRCAD FMUZR MBNFG
WGCEH VPIGK XCLFH NSHGB XWPRV SBXSR
JZUEJ WFAHQ KGNBS YOMBO ABYSR JHLNF
LCLFZ AZFOH GPNNL FOVYH SHZBX JQIEQ
WFMNW SBSGL ESUSW WFYVJ ZHUKP LCGBU
JCQFW GDWTV WQIAG VWPVV ACH (443)

38. VKSWD EXFCK CZKEX FTYZD IFNWA EUJTA
XRPCI MAXHG GRLNA VNQJY MWDWD GAVZW
DGIUS PVKWJ YYHYT LSNZZ EFVTK UHJTB
ZDIFN WAEHZ NXKAS HZL (108)

39. KOWYZ NMXHG HLNXB LGHAN RFOPD QYPNE
QWMEE FEFIG EEULJ LIQGA MRHVL RAWGZ
BNFXI UOMQX TETL (74)
[Probable word: PLEASANTON]

The following problems concern Repeating Key Systems with mixed cipher alphabets. Read the messages and recover the keywords used.

40. FQUHA WXDVI UWXCP HHVTP PQNNK RTNNX
DKHEQ KXZFN PQNYU OTSFQ UHAIW XHVPT
PZRXH VPXHV PBCZM GBSVM HKOIH PRKCK
JOWEM MBGVP PPRAC WDBXN QZHKJ PXPZO
LFOOI GVOXP VYDVR YAXTF GBFPN OPKYW
ULAEU SHQEP MQMYI MUOKW TFGQN LVEMM
CPFXH RULKG KLWXQ LBGPA GYUOW DEGBE
NGXPJ LJXOO IGVOX PEGBO RADIM EDLVP
BQNID KTBSG NTCPW KRIWP COHLA XFDCX
RALBP APZFN PQNNB GCMLR FSPWF GWGNB
XPWYF XZOLF MGIEU OWDEG BORNX PJWYE
UOXRR YBKAO WIEPH VNGXP VPBQN IFIYY
UVAYL XTBSE VPPNT PHRWM BERKH DFDHP
WNXPE KXPWP MENPR XDOBR MYIRF SPWFG
WVRNT KWLGG NTXVM OWDIY FJWMC XXFPX

QLBGP ACXNO WEWLH PDGTV MYIIJ KROPK
YWDLU RCLWE UYUKF FHWPQ LFPTB GCMLE
UKCPM MLQOY IENHX VWMXW WOUET PLIME
CSQYB ENPUL VPXSQ LGKRY B (561)

41. JUAAC HAXFR KKTUK YMSMU ZHUDI SFLUO
TCKQR RRUSW CEXZG NAKBU GEMHN IKQRP
IYKYC NTGRO QBEEJ WAKQH BSSJY ZJWAK
QHZYK PLUZC GB (102)

[This message has been enciphered using the same components as used in Problem 40 above.]

42. ADCMO GZRIT FUSOS WITZI UXFOR BZBMV
BUZCD XODCX PGJDY FAFDB BDF (53)

[The word CROSSROADS occurs in the message and the same cipher alphabet employed in Problem 40 above was used.]

The following problem concerns a Repeating Key System with a mixed plain component. Read the message and recover the keywords used.

43. SFYXF IOCOD OUXMC NCHEZ KPIIH SIGEM
ZYYMH PWEOT XKSMC XPTXH NCFSA EOKJO
TQMUM ZHWZO KTJEN AHLRD SXSVT DHAPL
LLGJE WOESB ZJTNM JNCXA SLROD STLIL
WSAXT EOMHG QCHSF LEVAI OUDOX LATIH
JPVDT OGXBC TAQJW DBYTM ZWPJD TWGAZ
ZILWS XBIYE MJAYX OEJQE VAIOY HWWSH
EUJEX VISBJ QYWXK FUFSA NSLHC ZLYEN
ITZLL TPCHG BTPWH QLAHT IHXSX OCJXF
YLLLG JEWCD ZUJRG RKTOJ ENAHL RDSXQ
MOFXF SSOCO PFWOI SLOBW ZTTIH QTLLV
WYFTJ ISJJM EUXSF AAXLI EMJOO AXSJL
JXJMU JQJSS VSFLJ PMHSL IBKWX PFQHI
ZEOOD MECKP UOTZL OMGCX ZRKTO YXFIO
WZGEV WXMFS BWWEC BJQWC STWKZ PJMXJ
UFNAH LDHAP LLLGJ EPJML WHXGA QPAHE
VAQLZ CEVCV UFQFV MVUIH GIWBS LGHGG

DLVAH GWMEB HAXBM SPYDX BQOFP EVAQL
ZCEVR EOGSI RCHBH ZAYAH VWIAX TEOMH
GWSLR NHBUY LREZA QLXSF AHIZK TPGEZ
RSVGA APCDR NIEVA PAPYL YAQTW SBKAF
LYDET SVKFP NZWLH PSQES ATOTY AUOOM
DMEAH NRDAX CVUUD EWHMM IZCSA XEBJQ
FCNTP LEXVE CNSFK AVJNC XNAHL NQGPE
UTYSH UAPAP NFESD ZPSQW XNVYF EVVMT
(750)

What are the keys on which the following equivalent primary components (secondary alphabets) are based:

44. N G S U H T R I V Y K W B L X C M Z D O J E P A F Q

45. O M D K U G N C J S Z R B I Q Y E A H P X V T F L W

46. J R H U F P Z M B E X K T I V G O C Q D S Y L A N W

47. Z Y A S D G K Q W C P N I E H M U X R T L B F J O V

Two messages, Message A (plaintext) and Message B (ciphertext), have been intercepted. It is suspected that Message A is the plaintext beginning of Message B (and only that portion of Message B that matches Message A is furnished). The enemy has been using a mixed sequence slid against itself. Determine the keyword upon which the primary component is based and the specific key used to encipher Message B.

48. Message A

WE ARE EXPECTING A MOVE TO BORTON SCHOOLHOUSE TONIGHT
SOON AFTER ONE AM TO DEFEND THE LINES EAST OF BORTON
SCHOOLHOUSE BE PREPARED AT THAT TIME TO MOVE OUT
PROMPTLY STOP OUR ADV....

Message B

CNUYW VLELY MXZCK AALLU PSSYU
PYVJM PACTV NUGDV KXWTU EASBK
XWGYV KNUUU PIGMW KIATW WXZLL

VRVDV XSSYU PYVJM PACTV NUGFW
SFGVK INNNU MVUDU JCGDV AALLV
NWEYV ADARE EWSVV NFUKC ....

The following problem concerns a Repeating Key System with mixed components. Read the message and recover the keywords used.

49. XJITZ FPVYV MNNTL CJIDT FNHTL XVWZT
HJOKH BZEYE VPNHZ NMAEA RWHXA RFWKK
VGAKH BSWMR DLCNC UAJEF QNSQC IVLJK
XZKNX CPXZX KLJLH RYQKE MBDXK HNFJE
AZUKH BSCZL BPNXD BAXMR BBHQC PPPCF
EJLGT GRXPE SBOLH HNVMO URGAV BFSPS
NWUZZ COLZP KJHPL JRKET HXTHR JWIDK
IITKH BSIZJ ANHAW NPJTE ABDXX JYFZO
ANKKK PAHYT TNNAL NVLPK CJOLH HNVMO
UCBKH BZHIZ DBDKH BSYES NBDXF PYZHT
TBLHC CWLRZ BNHXB BAHIZ BQGYW JWHCS
LIBXC PWLRZ BRHWH RYIEO MWBMS EVQML
ENUAW XWLRZ BCBTZ XYZWT FBPZL CYXVP
KBDXW XNJMJ KSTYW JWHIN RBDXH RYIZO
KNCHD PAIZN ALURL QKWMS MJCQA UYHTS
NNCDR XYAZE TRGMH RAHMR DTLKU BOAVW
YGLKO BARIL DQGRL YAHWK XJIAW WKBMT
BIGBW XVMMJ DMDAW QGIGW DUPXL XVMZE
YGBMU XYKEM KNHXW XYFCF NULPL CYBMJ
YGNXL YKCTS ALBNC NMCUE RJUQM YKCTS
VJSXC (605)

[Probable words: THE, COMMA, STOP, ARTILLERY, ENEMY, INFANTRY, POINT.]

The following two messages have been intercepted. Read the messages and reconstruct the alphabets and keywords used.

50\.

Message A

MUOUV DSWKN ICHGL BJSIM XOPJC IWNUR
MTOGG SDNOO IAHTP ZKKKE ONNVM GQOKJ

QCKAE YQQSO MOCBM HKJQC THSJJ OYWUY
HOJKN EJZJM LCZEO NNERJ OOMVI OHMQH
MCKGU JRICW NKOMY MMQHI YYUUF ICMKX
KEONN GZMJK NHYOH MRUFO PNRFT MIMMJ
DNORQ XJMXR QXAFM VECHT (200)

Message B

UQQCL OHTBP UAZFF FHDDJ KTOXF UCPQJ
UPQJF DWMQT UMZPU UCKGV QPMGU FVTCX
AIBDV SAZDT JQFAY MCXAI IMKXQ NSYQS
ZNXLM HQYXO SARVQ PMHOH QTJGD NWOZW
UIBJQ XOUAY MBCJS OJVZU SQQEX UAOCK
GVQPM TRJXL MWENW OEEXN UPEPS JROJX
WMQBZ KQJKB ZKPDY RVAZP (200)

The following message was intercepted shortly after the transmission of Messages A and B in Problem 50 above. It is believed that this message uses the same components as were used for Messages A and B. Read the message and determine its keyword.

51. WFKQF QRXLQ TFCCX GWELC PSAKW FAQRU
TFFAK ICCKG OCDKR EDJOQ PCWFK QFEXC
(60)

The following two messages were intercepted within minutes of each other. It is believed that the plaintext of both is the same. Read the messages, reconstruct the components used, and recover the keywords used.

52.

Message A

TBERJ SYQMI MREGJ HARBV UXJCF YEMEM
UTNCX IVSJE TBEBN KNPNV BSVPQ GTVBL
ABJRG YYGXD FZVRJ (75)

Message B

NQIPK DSFMT VFZZN TTEAG UIOJS PIBFV
WMNWU OHJNZ UHUVN RWSCF GLWZK STGHV
MQNPH GSSXP KDHNN (75)

As in the case of the two messages in Problem 52, the two messages in this problem were intercepted at approximately the same time, and their plaintext is believed to be the same. Read the messages, reconstruct the components used, and determine the keywords used.

53.

Message A

```
FUZYV TAQWF WDWUX QAZWL QUQTE NFALO
OPAKK MKWZD NKYFU MDTTG FFCAN NHPAO
TTPZK ODDXB IKZPU OXJTX (80)
```

Message B

```
UXAGT YEFLV BPEPT HZPOC LZJPE ULPJK
GRSCV FLTFL KFKXA YSJUX AHIMN UPYXK
DIOBV AUZUT JFUHA ZVAUX (80)
```

Read the following message. Determine the alphabets employed, the keys upon which they are based, and the specific-key used for the message's encipherment.

```
54. NCLOO ALTXJ ASNJQ SFBBL KHNUA HWWHP
    ULDGV UFJMB BPVST CQLKO PGIAZ NLYFR
    ZBLNS EZARP QFBHY BKPNW WQIWD NXQZF
    OYMGW QIIJN IRZBK AZXLO TVTXY CRYFZ
    MGIDG PZMFL QYQOS JOMLD UEVYY BMYNV
    YRMFA EFWQN GNCYC RYPND WBWUW GWXTC
    QWONW RHBKL JGDYE MUEQW QNLVZ WDPFM
    ESEJS BRVLG WMRLJ JJZAQ MVEEJ QIKWO
    BGOTL TCURZ BLZGG ZEFKW SXWQY MONGR
    SIWUP GDTMQ EGGKR TQLJJ MQEQW SXWQB
    RXSFT WIFES EJPBG NSXRZ BKAZX LOTVE
    SYCOL ZAYUW RHZVC KTWAK HMWHE FFVJA
    TVFAZ CBLTB RXSFT WIKLJ JRZBM ZDGRZ
    RMENR YTBCA NRRQN LVCMX WGLNO HDUNA
    VFGYM GLZGG ZSIWU PGWQN GEQGZ WRLPS
    EJWID NDVOT FMZYF XDYMQ NYZOM FAYXR
    WAIDN XRZBL ZGGZQ EWLHB JHHQU YFTSW
    MOIRF EFKPQ GAWIL JEYVR ZJVFU AOLZL
```

CYLBF SNDVO TFMZY FDMMQ VMGHS WMYNW
ZQFZG NXUTN LPQJA FFZLY FWMTV TYEDQ
UWGWZ TXQLH SUCWA MYNVA PNJND VOTFM
ZYFKA UKPHB JMNRL CRCAN ZLJJM OLZND
ZOHRR SNUQR LZQNX LDGQU GBYXB LOAWR
MFJLN PGWIM NDZOH RRSNU APNJP QWRVF
AZCBL TRLPS EJWID HYXRZ NSPQG IPFPX
NWRUW XTXGW MTVTY E (766)

[Probable words: STOP, FIRST, THE, ARTILLERY, DIVISION, AMBULANCES.]

The following message contains in its text a repetition of plaintext eighty nine letters long. Read the message, reconstruct the plain and cipher components, and determine the keywords used.

55. LIUDB USNYL BIDDK BZULA XMQZP QKUCW
SLCWL SLVUX IMMLP ANULP ANUWY INCZO
LQRDG YSHBS YNZPC SSPYY STWHO GBGMW
IBRID SLWHA TUPHK JODPW BKGZK JODPK
YLARN IBCUC AUWCY IYDLG IQNKE JTRPA
TUPIA TDNJK BRDDQ DBDDQ DBPKK DUUAY
ELHPF JMICF DJKWA JTSKA XEWAN YNELK
YLXXA TEWUO LTWHW CSKLP ANUHW CSKSH
(240)

COMPUTER PROGRAMS

The computer programs that follow have been especially written for this book by Wayne G. Barker. These computer programs, although written specifically in BASIC for use with a TRS-80 MODEL 4 computer, with usually minimum change are likewise applicable to other personal computers which use BASIC. The computer programs are only representative of the many programs that pertain to polyalphabetic substitution cipher systems. These type cipher systems particularly lend themselves to the use of computer programs to perform both cryptographic and cryptanalytic tasks. Indeed, virtually all the cryptanalytic operations described in this book can be duplicated and perhaps even improved upon by computer programs.

Each of the following listed computer programs is followed by a RUN of the program to show clearly the results obtained by using the program. The results of using the listed programs for the most part are self-evident. Only with respect to the fourth listed program, "Determining the Period of a Periodic Cipher", is an explanation of the program probably required. Using the program, the "period" of most simple polyalphabetic substitution ciphers can usually be easily discovered. The only qualification with respect to the program is that the "period" must be relatively short, up to about 25 letters, and the length of the ciphertext must be sufficient to provide a reasonable number of letters to be enciphered by each letter of the key. The computer program in turn "tests" various key-lengths, beginning first with a key-length of 1, then a key-length of 2, etc. For each key-length the program provides the average "index of coincidences" for the repetitions found in the various monoalphabetic distributions formed by the key-length being tested. In general, the closer that the average "IC" approaches .0667 (the probability of monographic coincidence in English telegraphic plaintext), the more likely it is that the key-length producing that average "IC" is the correct "period" or key-length of the ciphertext being examined. In this connection, it is advised that the student especially read Appendix 2, pages 108-118.

INDEX OF PROGRAMS

	Page
Vigenere Encipherment	148
True Beaufort Encipherment	150
Variant Beaufort Encipherment	152
Determining the Period of a Periodic Cipher	154
Vigenere Encipherment Using Mixed Alphabets	156

VIGENERE ENCIPHERMENT

```
10 REM -- "VIGENERE"
20 REM -- VIGENERE ENCIPHERMENT.
30 CLEAR 1000
40 DIM A$(250),K$(25),C(250),K(25),P(250),Q(250)
50 CLS
60 PRINT "Enter KEYWORD --"
70 INPUT K$
80 FOR I=1 TO LEN(K$)
90 K(I)=ASC(MID$(K$,I,1))
100 NEXT
110 PRINT
120 PRINT "Enter PLAINTEXT --"
130 INPUT A$
140 FOR I=1 TO LEN(A$)
150 Q(I)=ASC(MID$(A$,I,1))
160 NEXT
170 R=0
180 FOR S=1 TO LEN(A$)
190 IF Q(S)<65 OR Q(S)>90 THEN 220
200 R=R+1
210 P(R)=Q(S)
220 NEXT
230 PRINT
240 J=0
250 FOR I=1 TO R
260 J=J+1
270 IF J>LEN(K$) THEN J=0:GOTO 260
280 K=K(J)-65
290 IF P(I)>90-K THEN 320
300 C(I)=P(I)+K
310 GOTO 330
320 C(I)=(P(I)+K-90)+64
330 NEXT
340 PRINT "Ciphertext --"
350 PRINT
360 L=0
370 FOR M=1 TO 5
380 FOR N=1 TO 5
390 L=L+1
400 PRINT CHR$(C(L))" ";
410 IF L=R THEN 470
420 NEXT
430 PRINT "   ";
440 NEXT
450 PRINT
460 GOTO 370
470 PRINT:PRINT
480 PRINT "("R")"
490 END
```

```
Enter KEYWORD --
? BED

Enter PLAINTEXT --
? SEND SUPPLIES TO MORLEYS STATION

Ciphertext --

T I Q E W   X Q T O J   I V U S P   P V O F C   V T X D U
M R O

( 28 )
```

TRUE BEAUFORT ENCIPHERMENT

```
10 REM -- "BEAUFORT"
20 REM -- TRUE BEAUFORT ENCIPHERMENT.
30 CLEAR 1000
40 DIM A$(250),K$(25),C(250),K(25),P(250),Q(250)
50 CLS
60 PRINT "Enter KEYWORD --"
70 INPUT K$
80 FOR I=1 TO LEN(K$)
90 K(I)=ASC(MID$(K$,I,1))
100 NEXT
110 PRINT
120 PRINT "Enter PLAINTEXT --"
130 INPUT A$
140 FOR I=1 TO LEN(A$)
150 Q(I)=ASC(MID$(A$,I,1))
160 NEXT
170 R=0
180 FOR S=1 TO LEN(A$)
190 IF Q(S)<65 OR Q(S)>90 THEN 220
200 R=R+1
210 P(R)=Q(S)
220 NEXT
230 PRINT
240 J=0
250 FOR I=1 TO R
260 J=J+1
270 IF J>LEN(K$) THEN J=0:GOTO 260
280 K=K(J)-65
290 C(I)=P(I)+(26-2*(P(I)-K(J)))-K
300 IF C(I)>90 THEN C(I)=C(I)-26
310 NEXT
320 PRINT "Ciphertext --"
330 L=0
340 FOR M=1 TO 5
350 FOR N=1 TO 5
360 L=L+1
370 PRINT CHR$(C(L))" ";
380 IF L=R THEN 440
390 NEXT
400 PRINT "   ";
410 NEXT
420 PRINT
430 GOTO 340
440 PRINT:PRINT
450 PRINT "("R")"
460 END
```

```
Enter KEYWORD --
? COMET

Enter PLAINTEXT --
? SEND SUPPLIES

Ciphertext --
K K Z B B   I Z X T L   Y W

( 12 )
```

VARIANT BEAUFORT ENCIPHERMENT

```
10 REM -- "VARIANT"
20 REM -- VARIANT BEAUFORT ENCIPHERMENT.
30 CLEAR 1000
40 DIM A$(250),K$(25),C(250),K(25),P(250),Q(250)
50 CLS
60 PRINT "Enter KEYWORD --"
70 INPUT K$
80 FOR I=1 TO LEN(K$)
90 K(I)=ASC(MID$(K$,I,1))
100 NEXT
110 PRINT
120 PRINT "Enter PLAINTEXT --"
130 INPUT A$
140 FOR I=1 TO LEN(A$)
150 Q(I)=ASC(MID$(A$,I,1))
160 NEXT
170 R=0
180 FOR S=1 TO LEN(A$)
190 IF Q(S)<65 OR Q(S)>90 THEN 220
200 R=R+1
210 P(R)=Q(S)
220 NEXT
230 PRINT
240 J=0
250 FOR I=1 TO R
260 J=J+1
270 IF J>LEN(K$) THEN J=0:GOTO 260
280 K=K(J)-65
290 C(I)=P(I)-K
300 IF C(I)<65 THEN 320
310 GOTO 330
320 C(I)=C(I)+26
330 NEXT
340 PRINT "Ciphertext --"
350 PRINT
360 L=0
370 FOR M=1 TO 5
380 FOR N=1 TO 5
390 L=L+1
400 PRINT CHR$(C(L))" ";
410 IF L=R THEN 470
420 NEXT
430 PRINT "   ";
440 NEXT
450 PRINT
460 GOTO 370
470 PRINT:PRINT
480 PRINT "("R")"
490 END
```

```
Enter KEYWORD --
? COMET

Enter PLAINTEXT --
? SEND SUPPLIES

Ciphertext --

Q Q B Z Z   S B D H P   C E

( 12 )
```

DETERMINING THE PERIOD OF A PERIODIC CIPHER

```
10 REM -- "PERIOD"
20 REM -- DETERMINING THE PERIOD OF A PERIODIC CIPHER.
30 CLEAR 600
40 CLS
50 DIM C(255),Q(255),Z(90)
60 DEFINT I
70 INPUT "Test KEY LENGTHS to what length";K
80 PRINT
90 PRINT "Enter text of periodic cipher --"
100 INPUT A$
110 FOR I=1 TO LEN(A$)
120 Q(I)=ASC(MID$(A$,I,1))
130 NEXT
140 FOR S=1 TO LEN(A$)
150 IF Q(S)<65 OR Q(S)>90 THEN 180
160 R=R+1
170 C(R)=Q(S)
180 NEXT
190 PRINT
200 J=J+1
210 B=1
220 FOR I=1 TO R
230 N=N+1
240 IF N=B THEN Z(C(I))=Z(C(I))+1:T=T+1
250 IF N=J THEN N=0
260 NEXT
270 FOR I=65 TO 90
280 H=Z(I)*(Z(I)-1):M=M+H
290 Z(I)=0
300 NEXT
310 W=M/(T*(T-1)):A=A+W
320 H=0:M=0:N=0:T=0:W=0
330 IF B=J THEN 360
340 B=B+1
350 GOTO 220
360 PRINT "FOR KEY LENGTH "J "-- AVERAGE IC ="A/J
370 A=0
380 IF J=K THEN 400
390 GOTO 200
400 END
```

```
Test KEY LENGTHS to what length? 10

Enter text of periodic cipher --
? NFWWP NOMKI WPIDS CAAET QVZSE YOJSC AAAFG RVNHD
  WDSCA EGNFP FOEMT HXLJW PNOMK IQDBJ IVNHL TFNCS
  BGCRP

FOR KEY LENGTH  1 -- AVERAGE IC = .0372549
FOR KEY LENGTH  2 -- AVERAGE IC = .0363288
FOR KEY LENGTH  3 -- AVERAGE IC = .0283707
FOR KEY LENGTH  4 -- AVERAGE IC = .0392857
FOR KEY LENGTH  5 -- AVERAGE IC = .0426471
FOR KEY LENGTH  6 -- AVERAGE IC = .0285714
FOR KEY LENGTH  7 -- AVERAGE IC = .0932401
FOR KEY LENGTH  8 -- AVERAGE IC = .0522727
FOR KEY LENGTH  9 -- AVERAGE IC = .0320988
FOR KEY LENGTH  10 -- AVERAGE IC = .0253968
```

VIGENERE ENCIPHERMENT USING MIXED ALPHABETS

```
10 REM -- "MIXED"
20 REM -- VIGENERE ENCIPHERMENT USING MIXED ALPHABETS.
30 CLEAR 1000
40 CLS
50 DIM B(26,90),C(255),K(25),N(255),P(255),Q(255)
60 DIM CA(26),CC(26),PA(26),PC(26),PD(26)
70 PRINT "Enter KEYWORD --"
80 INPUT K$
90 FOR I=1 TO LEN(K$)
100 K(I)=ASC(MID$(K$,I,1))
110 NEXT
120 PRINT
130 PRINT "Enter INDEX LETTER --"
140 INPUT E$
150 PRINT
160 PRINT "Enter PLAINTEXT COMPONENT --"
170 INPUT PC$
180 IF LEN(PC$)<>26 THEN 160
190 FOR I=1 TO 26
200 PC(I)=ASC(MID$(PC$,I,1))
210 NEXT
220 PRINT
230 PRINT "Enter CIPHERTEXT COMPONENT --"
240 INPUT CC$
250 IF LEN(CC$)<>26 THEN 230
260 FOR I=1 TO 26
270 CC(I)=ASC(MID$(CC$,I,1))
280 NEXT
290 PRINT
300 PRINT "Enter PLAINTEXT --"
310 INPUT P$
320 FOR I=1 TO LEN(P$)
330 Q(I)=ASC(MID$(P$,I,1))
340 NEXT
350 R=0
360 FOR I=1 TO LEN(P$)
370 IF Q(I)<65 OR Q(I)>90 THEN 400
380 R=R+1
390 P(R)=Q(I)
400 NEXT
410 FOR I=1 TO 26
420 IF PC(I)=ASC(E$) THEN X=I:GOTO 440
430 NEXT
440 FOR I=1 TO 26
450 G=X+S
460 IF G>26 THEN G=G-26
470 PA(I)=PC(G)
480 S=S+1
490 NEXT
500 J=0
```

```
510 J=J+1
520 FOR I=1 TO 26
530 IF K(J)=CC(I) THEN Y=I:T=0:GOTO 560
540 NEXT
550 T=0
560 FOR I=1 TO 26
570 H=Y+T
580 IF H>26 THEN H=H-26
590 CA(I)=CC(H)
600 H=0
610 T=T+1
620 NEXT
630 FOR I=1 TO 26
640 B(J,I)=CA(I)
650 NEXT
660 IF J=LEN(K$) THEN 680
670 GOTO 510
680 I=0
690 I=I+1
700 FOR J=65 TO 90
710 IF PA(I)=J THEN PD(J-64)=I: GOTO 730
720 NEXT
730 IF I=26 GOTO 750
740 GOTO 690
750 I=0
760 PRINT
770 PRINT "CIPHERTEXT:"
780 PRINT
790 FOR K=1 TO LEN(K$)
800 W=W+1
810 L=P(W)-64
820 M=PD(L)
830 N(W)=B(K,M)
840 IF W=R THEN 870
850 NEXT
860 GOTO 790
870 W=0
880 FOR M=1 TO 5
890 FOR N=1 TO 5
900 W=W+1
910 PRINT CHR$(N(W))" ";
920 IF W=R THEN 980
930 NEXT
940 PRINT "   ";
950 NEXT
960 PRINT
970 GOTO 880
980 PRINT
990 PRINT "("R")"
1000 PRINT
1010 END
```

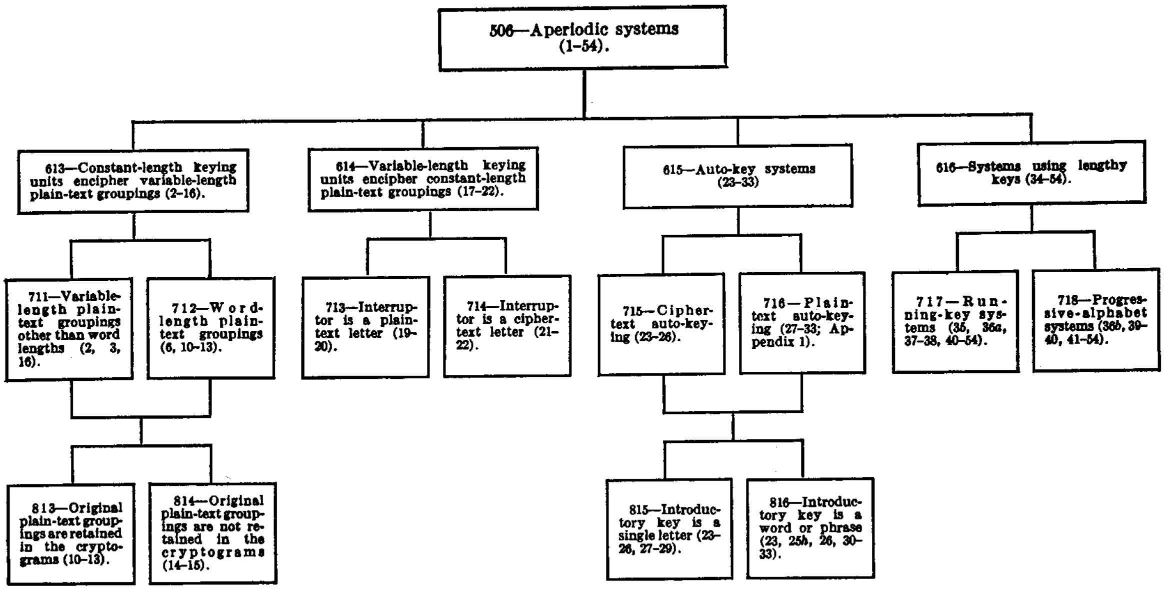
```
Enter KEYWORD ==
? JOURNEY

Enter INDEX LETTER ==
? Q

Enter PLAINTEXT COMPONENT ==
? QUESTIONABLYCDFGHJKMPRVWXZ

Enter CIPHERTEXT COMPONENT ==
? QUESTIONABLYCDFGHJKMPRVWXZ

Enter PLAINTEXT ==
? HAVE DIRECTED SECOND REGIMENT TO CONDUCT THORO
  RECONNAISSAINCE IN THE DIRECTION OF HORSESHOE
  FALLS

CIPHERTEXT:

N F W W P   N O M K I   W P I D S   C A A E T   Q V Z S E
Y O J S C   A A A F G   R V N H D   W D S C A   E G N F P
F O E M T   H X L J W   P N O M K   I Q D B J   I V N H L
T F N C S   B G C R P
( 85 )
```

A CRYPTOGRAPHIC SERIES

(60)

MILITARY CRYPTANALYSIS
PART III

Simpler Varieties of Aperiodic Substitution Systems


506—Aperiodic systems (1–54).
613—Constant-length keying units encipher variable-length plain-text groupings (2–16).
614—Variable-length keying units encipher constant-length plain-text groupings (17–22).
615—Auto-key systems (23–33)
616—Systems using lengthy keys (34–54).
711—Variable-length plain-text groupings other than word lengths (2, 3, 16).
712—Word-length plain-text groupings (6, 10–13).
713—Interruptor is a plain-text letter (19–20).
714—Interruptor is a cipher-text letter (21–22).
715—Cipher-text auto-keying (23–26).
716—Plain-text auto-keying (27–33; Appendix 1).
717—Running-key systems (35, 36a, 37–38, 40–54).
718—Progressive-alphabet systems (36b, 39–40, 41–54).
813—Original plain-text groupings are retained in the cryptograms (10–13).
814—Original plain-text groupings are not retained in the cryptograms (14–15).
815—Introductory key is a single letter (23–26, 27–29).
816—Introductory key is a word or phrase (23, 25b, 26, 30–33).

[Numbers in parentheses refer to Paragraph Numbers in this text]

by

William F. Friedman

From Aegean Park Press

- **MILITARY CRYPTANALYSIS PART III** **1**
  - SECTION I 5
    - INTRODUCTORY 5
  - SECTION II 9
    - SOLUTION OF SYSTEMS USING CONSTANT-LENGTH 9
  - SECTION III 12
    - SOLUTION OF SYSTEMS USING CONSTANT-LENGTH 12
  - SECTION IV 18
    - SOLUTION OF SYSTEMS USING CONSTANT-LENGTH 18
  - SECTION V 23
    - SOLUTION OF SYSTEMS USING VARIABLE-LENGTH 23
  - SECTION VI 32
    - REVIEW OF AUTO-KEY SYSTEMS 32
  - SECTION VII 34
    - SOLUTION OF CIPHER-TEXT AUTO-KEY SYSTEMS 34
  - SECTION VIII 49
    - SOLUTION OF PLAIN-TEXT AUTO-KEY SYSTEMS 49
  - SECTION IX 54
    - METHODS OF LENGTHENING OR EXTENDING THE KEY 54
  - SECTION X 57
    - GENERAL PRINCIPLES UNDERLYING SOLUTION OF 57
  - SECTION XI 62
    - THE COINCIDENCE OR x TEST 62
  - SECTION XII 77
    - THE CROSS-PRODUCT SUM OR x TEST 77
  - SECTION XIII 83
    - APPLYING THE CROSS-PRODUCT SUM OR x TEST 83
  - SECTION XIV 98
    - THE MONOALPHABETICITY OR + TEST 98
  - SECTION XV 101
    - CONCLUDING REMARKS 101
  - APPENDIX 1 102
  - INDEX 121
    - Analytical Key for Military Cryptanalysis, Part III 123
  - BOOKS IN THE CRYPTOGRAPHIC SERIES 124

MILITARY CRYPTANALYSIS
Part III

SIMPLER VARIETIES
OF APERIODIC SUBSTITUTION SYSTEMS

By

WILLIAM F. FRIEDMAN

This is a quality reproduction of a U.S. Military text, originally published in 1939 — declassified from CONFIDENTIAL in December 1992.

ISBN: 0-89412-196-0 (soft cover)

AEGEAN PARK PRESS
P.O. Box 2837
Laguna Hills, California 92654
(714) 586-8811
FAX (714) 586-8269

Manufactured in the United States of America

MILITARY CRYPTANALYSIS. PART III. SIMPLER VARIETIES OF APERIODIC SUBSTITUTION SYSTEMS

CONTENTS

Section		Paragraphs	Pages
I.	Introductory	1–4	1–4
II.	Solution of systems using constant-length keying units to encipher variable-length plain-text groupings, I	5–9	5–7
III.	Solution of systems using constant-length keying units to encipher variable-length plain-text groupings, II	10–13	8–13
IV.	Solution of systems using constant-length keying units to encipher variable-length plain-text groupings, III	14–16	14–18
V.	Solution of systems using variable-length keying units to encipher constant-length plain-text groupings	17–22	19–27
VI.	Review of auto-key systems	23	28–29
VII.	Solution of cipher-text auto-key systems	24–29	30–43
VIII.	Solution of plain-text auto-key systems	30–33	45–49
IX.	Methods of lengthening or extending the key	34–36	50–52
X.	General principles underlying solution of systems employing long or continuous keys	37–40	53–57
XI.	The "coincidence" or "κ test"	41–44	58–72
XII.	The "cross-product" sum or "χ test"	45–48	73–78
XIII.	Applying the cross-product sum or χ test	49–51	79–93
XIV.	The "monoalphabeticity" or "Φ test"	52–54	94–96
XV.	Concluding remarks	55–56	97
Appendix 1.	Additional notes on methods for solving plain-text auto-keyed ciphers	1–7	98–116
Index			117–118
Analytical key for Part III			119

SECTION I

INTRODUCTORY

Paragraph
Preliminary remarks 1
General remarks upon the nature of cryptographic periodicity 2
Effects of varying the length of the plain-text groupings 3
Primary and secondary periods; resultant periods 4

1. **Preliminary remarks.**—*a.* The text immediately preceding this devoted itself almost exclusively to polyalphabetic substitution systems of the type called repeating-key ciphers. It was seen how a regularity in the employment of a limited number of alphabets results in the manifestation of periodicity or cyclic phenomena in the cryptogram, by means of which the latter may be solved. The difficulty in solution is directly correlated with the type and number of cipher alphabets employed in specific examples.

b. Two procedures suggest themselves for consideration when the student cryptanalyst realizes the foregoing circumstances and thinks of methods to eliminate the weaknesses inherent in this cryptographic system. First, noting that the difficulties in solution increase as the length of the key increases, he may study the effects of employing much longer keys to see if one would be warranted in placing much trust in that method of increasing the security of the messages. Upon second thought, however, remembering that as a general rule the first step in the solution consists in ascertaining the number of alphabets employed, it seems to him that the most logical thing to do would be to use a procedure which will avoid periodicity altogether, will thus eliminate the cyclic phenomena that are normally manifested in cryptograms of a periodic construction, and thus prevent an enemy cryptanalyst from taking even a first step toward solution. In other words, he will investigate the possibilities of *aperiodic* systems first and if the results are unsatisfactory, he will then see what he can do with systems using lengthy keys.

c. Accordingly, the first part of this text will be devoted to an examination of certain of the very simple varieties of aperiodic, polyalphabetic substitution systems; after this, methods of extending or lengthening short mnemonic keys, and systems using lengthy keys will be studied.

2. **General remarks upon the nature of cryptographic periodicity.**—*a.* When the thoughtful student considers the matter of periodicity in polyalphabetic substitution systems and tries to ascertain its real nature, he notes, with some degree of interest and surprise perhaps that it is composed of *two* fundamental factors, because there are in reality *two* elements involved in its production. He has, of course, become quite familiar with the idea that periodicity necessitates the use of a keying element and that the latter must be employed in a cyclic manner. But he now begins to realize that there is another element involved, the significance of which he has perhaps not fully appreciated, *viz*, that unless the key is applied to constant-length plain-text groups no periodicity will be manifested externally by the cryptogram, despite the repetitive or cyclic use of a constant-length key. This realization is quickly followed by the idea that possibly all periodicity may be avoided or suppressed by either or both of two ways: (1) By using constant-length keying units to encipher variable-length plain-text groupings or (2) by using variable-length keying units to encipher constant-length plain-text groupings.

b. The student at once realizes also that the periodicity exhibited by repeating-key ciphers of the type studied in the preceding text is of a very simple character. There, successive letters of the repetitive key were applied to successive letters of the text. In respect to the employment of the key, the cryptographic or keying process may be said to be *constant* or *fixed* in character. This terminology remains true even if a single keying unit serves to encipher two or more letters at a time, provided only that the groupings of plain-text letters are constant in length. For example, a single key letter may serve to encipher two successive plain-text letters; if the key is repetitive in character and the message is sufficient in length, periodicity will still be manifested by the cryptogram and the latter can be solved by the methods indicated in the preceding text.[1] Naturally, those methods would have to be modified in accordance with the specific type of grouping involved. In this case the factoring process would disclose an apparent key length twice that of the real length. But study of the frequency distributions would soon show that the 1st and 2d distributions were similar, the 3d and 4th, the 5th and 6th, and so on, depending upon the length of the key. The logical step is therefore to combine the distributions in proper pairs and proceed as usual.

c. In all such cases of encipherment by constant-length groupings, the apparent length of the period (as found by applying the factoring process to the cryptogram) is a multiple of the real length and the multiple corresponds to the length of the groupings, that is, the number of plain-text letters enciphered by the same key letter.

d. The point to be noted, however, is that all these cases are still periodic in character, because *both* the keying units and the plain-text groupings are constant in length.

3. Effects of varying the length of the plain-text groupings.—*a.* But now consider the effects of making one or the other of these two elements *variable* in length. Suppose that the plain-text groupings are made variable in length and that the keying units are kept constant in length. Then, even though the key may be cyclic in character and may repeat itself many times in the course of encipherment, external periodicity is suppressed, *unless the law governing the variation in plain-text groupings is itself cyclic in character, and the length of the message is at least two or more times that of the cycle applicable to this variable grouping.*

b. (1) For example, suppose the correspondents agree to use reversed standard cipher alphabets with the key word SIGNAL, to encipher a message, the latter being divided up into groups as shown below:

S	I	G	N	A	L	S	I	G	N	A	L	S	I	G
1	12	123	1234	12345	1	12	123	1234	12345	1	12	123	1234	12345
C	OM	MAN	DING	GENER	A	LF	IRS	TARM	YHASI	S	SU	EDO	RDER	SEFFE
Q	UW	UGT	KFAH	UWNWJ	L	HN	ARQ	NGPU	PGNVF	I	TR	OPE	RFER	OCBBC

N	A	L	S	I	G	N	A	L	S	I	G	N	A	L
1	12	123	1234	12345	1	12	123	1234	12345	1	12	123	1234	12345
C	TI	VET	WENT	YFIRS	T	AT	NOO	NDIR	ECTIN	G	TH	ATT	ELEP	HONES
L	HS	QHS	WOFZ	KDARQ	N	NU	NMM	YIDU	OQZKF	C	NZ	NUU	WPWL	EXYHT

S	I	G	N	A	L	S	I
1	12	123	1234	12345	1	12	123
C	OM	MAS	WITC	HBOAR	D	SC	OMM...
Q	UW	UGO	RFUL	TZMAJ	I	AQ	UWW...

CRYPTOGRAM

QUWUG	TKFAH	UWNWJ	LHNAR	QNGPU	PGNVF	ITROP	ERFER
OCBBC	LHSQH	SWOFZ	KDARQ	NNUNM	MYIDU	OQZKF	CNZNU
UWPWL	EXYHT	QUWUG	ORFUL	TZMAJ	IAQUW	W...	

FIGURE 1.

---

[1] In this connection, see Section III, *Military Cryptanalysis, Part II*,

(2) The cipher text in this example (Fig. 1) shows a tetragraphic and a pentagraphic repetition. The two occurrences of QUWUG (=COMMA) are separated by an interval of 90 letters; the two occurrences of ARQN (=IRST) by 39 letters. The former repetition (QUWUG), it will be noted, is a true periodic repetition, since the plain-text letters, their grouping, and the key letters are identical. The interval in this case, if counted in terms of letters, is the product of the keying cycle, 6, by the grouping cycle, 15. The latter repetition (ARQN) is not a true periodic repetition in the sense that both cycles have been completed at the same point, as is the case in the former repetition. It is true that the cipher letters ARQN, representing IRST both times, are produced by the same key letters, I and G, but the enciphering points in the grouping cycle are different in the two cases. Repetitions of this type may be termed *partially periodic* repetitions, to distinguish them from those of the *completely periodic* type.

c. When the intervals between the two repetitions noted above are more carefully studied, especially from the point of view of the interacting cycles which brought them about, it will be seen that counting according to *groupings* and not according to single letters, the two pentagraphs QUWUG are separated by an interval of 30 groupings. Or, if one prefers to look at the matter in the light of the keying cycle, the two occurrences of QUWUG are separated by 30 key letters. Since the key is but 6 letters in length, this means that the key has gone through 5 cycles. Thus, the number 30 is the product of the number of letters in the keying cycle (6) by the number of different-length groupings in the grouping cycle (5). The interaction of these two cycles may be conceived of as partaking of the nature of two gears which are in mesh, one driven by the other. One of these gears has 6 teeth, the other 5, and the teeth are numbered. If the two gears are adjusted so that the "number 1 teeth" are adjacent to each other, and the gears are caused to revolve, these two teeth will not come together again until the larger gear has made 5 revolutions and the smaller one 6. During this time, a total of 30 meshings of individual teeth will have occurred. But since one revolution of the smaller gear (=the grouping cycle) represents the encipherment of 15 letters, when translated in terms of letters, the 6 complete revolutions of this gear mean the encipherment of 90 letters. This accounts for the period of 90, when stated in terms of letters.

d. The two occurrences of the other repetition, ARQN, are at an interval of 39 *letters;* but in terms of the number of intervening groupings, the interval is 12, which is obviously two times the length of the keying cycle. In other words, the key has in *this* case passed through 2 cycles.

e. In a long message enciphered according to such a scheme as the foregoing there would be many repetitions of both types discussed above (the completely periodic and the partially periodic) so that the cryptanalyst might encounter some difficulty in his attempts to reach a solution, especially if he had no information as to the basic system. It is to be noted in this connection that if any one of the groupings exceeds say 5, 6, or 7 letters in length, the scheme may give itself away rather easily, since it is clear that *within each grouping the encipherment is strictly monoalphabetic.* Therefore, in the event of groupings of more than 5 or 6 letters, the monoalphabetic equivalents of tell-tale words such as ATTACK, BATTALION, DIVISION, etc., would stand out. The system is most efficacious, therefore, with short groupings.

f. It should also be noted that there is nothing about the scheme which requires a regularity in the grouping cycle such as that embodied in the example. A lengthy grouping cycle such as the one shown below may just as easily be employed, it being guided by a key of its own; for example, the *number* of dots and dashes contained in the International Morse signals for the letters composing the phrase DECLARATION OF INDEPENDENCE might be used. Thus, A (.—) has 2, B (—...) has 4, and so on. Hence:

D	E	C	L	A	R	A	T	I	O	N	O	F	I	N	D	E	P	E	N	D	E	N	C	E
3	1	4	4	2	3	2	1	2	3	2	3	4	2	2	3	1	4	1	2	3	1	2	4	1

The grouping cycle is 3+1+4+4+2 . . ., or 60 letters in length. Suppose the same phrase is used as an enciphering key for determining the selection of cipher alphabets. Since the phrase contains 25 letters, the complete period of the system would be the least common multiple of 25 and 60 or 300 letters. This system might appear to yield a very high degree of cryptographic security. But the student will see as he progresses that the security is not so high as he may at first glance suppose it to be.

4. Primary and secondary periods; resultant periods.—*a.* It has been noted that the length of the complete period in a system such as the foregoing is the least common multiple of the length of the two component or interacting periods. In a way, therefore, since the *component* periods constitute the *basic* element of the scheme, they may be designated as the *basic* or *primary* periods. These are also *hidden* or *latent* periods. The *apparent* or *patent* period, that is, the complete period, may be designated as the *secondary* or *resultant* period. In certain types of cipher machines there may be more than two primary periods which interact to produce a resultant period; also, there are cases in which the latter may interact with another primary period to produce a tertiary period; and so on. The *final*, or *resultant*, or *apparent* period is the one which is usually ascertained first as a result of the study of the intervals between repetitions. This may or may not be broken down into its component primary periods.

b. Although a solution may often be obtained without breaking down a resultant period into its component primary periods, the reading of many messages pertaining to a widespread system of secret communication is much facilitated when the analysis is pushed to its lowest level, that is, to the point where the final cryptographic scheme has been reduced to its simplest terms. This may involve the discovery of a multiplicity of simple elements which interact in successive cryptographic strata.

SECTION II

SOLUTION OF SYSTEMS USING CONSTANT-LENGTH KEYING UNITS TO ENCIPHER VARIABLE-LENGTH PLAIN-TEXT GROUPINGS, I

	Paragraph
Introductory remarks	5
Aperiodic encipherment produced by groupings according to word lengths	6
Solution when direct standard cipher alphabets are employed	7
Solution when reversed standard cipher alphabets are employed	8
Comments on foregoing cases	9

5. Introductory remarks.—*a.* The system described in paragraph 3 above is obviously not to be classified as aperiodic in nature, despite the injection of a variable factor which in that case was based upon irregularity in the length of one of the two elements involved in polyalphabetic substitution. The variable factor was there subject to a law which in itself was periodic in character.

b. To make such a system truly aperiodic in character, by elaborating upon the basic scheme for producing variable-length plain-text groupings, would be possible, but impractical. For example, using the same method as is given in paragraph 3*f* for determining the lengths of the groupings, one might employ the text of a book; and if the latter is longer than the message to be enciphered, the cryptogram would certainly show no periodicity as regards the intervals between repetitions, which would be plentiful. However, as already indicated, such a scheme would not be very practical for regular communication between a large number of correspondents, for reasons which are no doubt apparent. The book would have to be safeguarded as would a code; enciphering and deciphering would be quite slow, cumbersome, and subject to error; and, unless the same key text were used for all messages, methods or indicators would have to be adopted to show exactly where encipherment begins in each message. A simpler method for producing constantly changing, aperiodic plain-text groupings therefore, is to be sought.

6. Aperiodic encipherment produced by groupings according to word lengths.—*a.* The simplest method for producing aperiodic plain-text groupings is one which has doubtless long ago presented itself to the student, *viz.*, encipherment according to the actual word lengths of the message to be enciphered.

b. Although the *average* number of letters composing the words of any alphabetical language is fairly constant, *successive* words comprising plain text vary a great deal in this respect, and this variation is subject to no law.[1] In telegraphic English, for example, the mean length of words is 5.2 letters; the words may contain from 1 to 15 or more letters, but the successive words vary in length in an extremely irregular manner, no matter how long the text may be.

c. As a consequence, the use of word lengths for determining the number of letters to be enciphered by each key letter of a repetitive key commends itself to the inexperienced cryptographer as soon as he comes to understand the way in which repeating-key ciphers are solved. If there is no periodicity in the cryptograms, how can the letters of the cipher text, written in

---

[1] It is true, of course, that the differences between two writers in respect to the lengths and characters of the words contained in their personal vocabularies are often marked and can be measured. These differences may be subject to certain laws, but the latter are not of the type in which we are interested, being psychological rather than mathematical in character. See Rickert, E., *New Methods for the Study of Literature*, University of Chicago Press, Chicago, 1927.

5-letter groups, be distributed into their respective monoalphabets? And if this very first step is impossible, how can the cryptograms be solved?

7. **Solution when direct standard cipher alphabets are employed.**—*a.* Despite the foregoing rhetorical questions, the solution of this case is really quite simple. It merely involves a modification of the method given in a previous text,[2] wherein solution of a monoalphabetic cipher employing a direct standard alphabet is accomplished by completing the plain-component sequence. There, all the words of the entire message come out on a single generatrix of the completion diagram. In the present case, since the individual, separate words of a message are enciphered by different key letters, *these words will reappear on different generatrices of the diagram.* All the cryptanalyst has to do is to pick them out. He can do this once he has found a good starting point, by using a little imagination and following clues afforded by the context.

b. An example will make the method clear. The following message (note its brevity) has been intercepted:

```
TRECS  YGETI  LUVWV  IKMQI  RXSPJ
SVAGR  XUXPW  VMTUC  SYXGX  VHFFB  LLBHG
```

c. Submitting the message to routine study, the first step is to use normal alphabet strips and try out the possibility of direct standard alphabets having been used. The completion diagram for the first 10 letters of the message is shown in figure 2.

d. Despite the fact that the text does not all reappear on the same generatrix, the solution is a very simple matter because the first three words of the message are easily found: CAN YOU GET. The key letters may be sought in the usual manner and are found to be REA. One may proceed to set up the remaining letters of the message on sliding normal alphabets, or one may assume various keywords such as READ, REAL, REAM, etc., and try to continue the decipherment in that way. The former method is easier. The completed solution is as follows:

R	E	A	D	E	R	S
CAN	YOU	GET	FIRST	REGIMENT	BY	RADIO
TRE	CSY	GET	ILUVW	VIKMQIRX	SP	JSVAG

D	I	G	E	S	T
OUR	PHONE	NOW	OUT	OF	COMMISSION
RXU	XPWVM	TUC	SYX	GX	VHFFBLLBHG

e. Note the key in the foregoing case: It is composed of the successive key letters of the phrase READERS DIGEST.

f. The only difficult part of such a solution is that of making the first step and getting a start on a word. If the words are short it is rather easy to overlook good possibilities and thus spend some time in fruitless searching. However, solution must come; if nothing good appears at the beginning of the message, search should be made in the interior of the cryptogram or at the end.

T	R	E	C	S	Y	G	E	T	I
U	S	F	D	T	Z	H	F	U	J
V	T	G	E	U	A	I	G	V	K
W	U	H	F	V	B	J	H	W	L
X	V	I	G	W	C	K	I	X	M
Y	W	J	H	X	D	L	J	Y	N
Z	X	K	I	Y	E	M	K	Z	O
A	Y	L	J	Z	F	N	L	A	P
B	Z	M	K	A	G	O	M	B	Q
C	A	N	L	B	H	P	N	C	R
D	B	O	M	C	I	Q	O	D	S
E	C	P	N	D	J	R	P	E	T
F	D	Q	O	E	K	S	Q	F	U
G	E	R	P	F	L	T	R	G	V
H	F	S	Q	G	M	U	S	H	W
I	G	T	R	H	N	V	T	I	X
J	H	U	S	I	O	W	U	J	Y
K	I	V	T	J	P	X	V	K	Z
L	J	W	U	K	Q	Y	W	L	A
M	K	X	V	L	R	Z	X	M	B
N	L	Y	W	M	S	A	Y	N	C
O	M	Z	X	N	T	B	Z	O	D
P	N	A	Y	O	U	C	A	P	E
Q	O	B	Z	P	V	D	B	Q	F
R	P	C	A	Q	W	E	C	R	G
S	Q	D	B	R	X	F	D	S	H

FIGURE 2.

---

[2] *Military Cryptanalysis, Part I,* Par. 20.

8. Solution when reversed standard cipher alphabets are employed.—It should by this time hardly be necessary to indicate that the only change in the procedure set forth in paragraph 7*c*, *d* in the case of reversed standard cipher alphabets is that the letters of the cryptogram must be converted into their plain-component (direct standard) equivalents before the completion sequence is applied to the message.

9. Comments on foregoing cases.—*a*. The foregoing cases are so simple in nature that the detailed treatment accorded them would seem hardly to be warranted at this stage of study. However, they are necessary and valuable as an introduction to the more complicated cases to follow.

b. Throughout this text, whenever encipherment processes are under discussion, the pair of enciphering equations commonly referred to as characterizing the so-called Vigenère method will be understood, unless otherwise indicated. This method involves the pair of enciphering equations $\Theta_{i/1}=\Theta_{k/2}$; $\Theta_{p/1}=\Theta_{c/2}$, that is, the index letter, which is usually the initial letter of the plain component, is set opposite the key letter on the cipher component; the plain-text letter to be enciphered is sought on the plain component and its equivalent is the letter opposite it on the cipher component.[3]

c. The solution of messages prepared according to the two preceding methods is particularly easy, for the reason that standard cipher alphabets are employed and these, of course, are derived from *known* components. The significance of this statement should by this time be quite obvious to the student. But what if mixed alphabets are employed, so that one or both of the components upon which the cipher alphabets are based are unknown sequences? The simple procedure of completing the plain component obviously cannot be used. Since the messages are polyalphabetic in character, and since the process of factoring cannot be applied, it would seem that the solution of messages enciphered in different alphabets and according to word lengths would be a rather difficult matter. However, it will soon be made clear that the solution is not nearly so difficult as first impression might lead the student to imagine.

---

[3] See in this connection, *Military Cryptanalysis, Part II*, Section II, and Appendix 1.

SECTION III

SOLUTION OF SYSTEMS USING CONSTANT-LENGTH KEYING UNITS TO ENCIPHER VARIABLE-LENGTH PLAIN-TEXT GROUPINGS, II

	Paragraph
Solution when the original word lengths are retained in the cryptogram	10
Solution when other types of alphabets are employed	11
Isomorphism and its importance in cryptanalytics	12
Illustration of the application of phenomena of isomorphism in solving a cryptogram	13

10. Solution when the original word lengths are retained in the cryptogram.—*a.* This case will be discussed not because it is encountered in practical military cryptography but because it affords a good introduction to the case in which the original word lengths are no longer in evidence in the cryptogram, the latter appearing in the usual 5-letter groups.

b. Reference is made at this point to the phenomenon called idiomorphism, and its value in connection with the application of the principles of solution by the "probable-word" method, as explained in a previous text.[1] When the original word lengths of a message are retained in the cryptogram, there is no difficulty in searching for and locating idiomorphs and then making comparisons between these idiomorphic sequences in the message and special word patterns set forth in lists maintained for the purpose. For example, in the following message note the underlined groups and study the letters within these groups:

MESSAGE

XIXLP EQVIB VEFHAPFVT RT XWK PWEWIWRD XM
NTJCTYZL OAS XYQ ARVVRKFONT BH SFJDUUXFP
OUVIGJPF ULBFZ RV DKUKW ROHROZ

IDIOMORPHIC SEQUENCES

(1) PWEWIWRD (2) ARVVRKFONT (3) SFJDUUXFP

(4) ROHROZ

c. Reference to lists of words commonly found in military text and arranged according to their idiomorphic patterns or formulae soon gives suggestions for these cipher groups. Thus:

(1)	PWEWIWRD DIVISION	(3)	SFJDUUXFP ARTILLERY
(2)	ARVVRKFONT BATTALIONS	(4)	ROHROZ OCLOCK

---

[1] *Military Cryptanalysis, Part I,* Par. 33 *a–d*, inclusive.

d. With these assumed equivalents a reconstruction skeleton or diagram of cipher alphabets (forming a portion of a quadricular table) is established, on the hypothesis that the cipher alphabets have been derived from the sliding of a mixed component against the normal sequence. First it is noted that since $O_p=R_c$ both in the word DIVISION and in the word OCLOCK their cipher equivalents must be in the same alphabet. The reconstruction skeleton is then as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Division, o'clock.....(1)			O	P					W		Z	H		D	R				I			E				
Battalion ....(2)	R	A							F			K		N	O				T	V						
Artillery .....(3)	S				X				D			U						F		J					P	

FIGURE 3*a*.

e. Noting that the interval between O and R in the first and second alphabets is the same, direct symmetry of position is assumed In a few moments the first alphabet in the skeleton becomes as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
(1)		N	O	P		S	T	V	W	X	Z	H		D	R	A	U		I			E	F		J	K
(2)	R	A							F			K		N	O				T	V						
(3)	S				X				D			U						F		J					P	

FIGURE 3*b*.

f. The key word upon which the mixed component is based is now not difficult to find: HYDRAULIC.

g. (1) To decipher the entire message, the simplest procedure is to convert the cipher letters into their plain-component equivalents (setting the HYDRAULIC . . . Z sequence against the normal alphabet at any point of coincidence) and then completing the plain-component sequence, as usual. The words of the message will then reappear on different generatrices. The key letters may then be ascertained and the solution completed. Thus, for the first three words, the diagram is as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z

X	I	X	L	P
Y	H	Y	G	S
Z	I	Z	H	T
A	J	A	I	U
B	K	B	J	V
C	L	C	K	W
D	M	D	L	X
E	N	E	M	Y

$A_p = S_c$

E	Q	V	I	B
K	T	W	H	J
L	U	X	I	K
M	V	Y	J	L
N	W	Z	K	M
O	X	A	L	N
P	Y	B	M	O
Q	Z	C	N	P
R	A	D	O	Q
S	B	E	P	R
T	C	F	Q	S
U	D	G	R	T
V	E	H	S	U
W	F	I	T	V
X	G	J	U	W
Y	H	K	V	X
Z	I	L	W	Y
A	J	M	X	Z
B	K	N	Y	A
C	L	O	Z	B
D	M	P	A	C
E	N	Q	B	D
F	O	R	C	E

$A_p = U_c$

V	E	F	H	A	P	F	V	T
W	K	L	A	E	S	L	W	V
X	L	M	B	F	T	M	X	W
Y	M	N	C	G	U	N	Y	X
Z	N	O	D	H	V	O	Z	Y
A	O	P	E	I	W	P	A	Z
B	P	Q	F	J	X	Q	B	A
C	Q	R	G	K	Y	R	C	B
D	R	S	H	L	Z	S	D	C
E	S	T	I	M	A	T	E	D

$A_p = P_c$

Figure 4.

(2) The key for the message is found to be SUPREME COURT and the complete message is as follows:

Solution

S	U	P	R	E	M	E	C	O	U
ENEMY	FORCE	ESTIMATED	AS	ONE	DIVISION	OF	INFANTRY	AND	TWO
XIXLP	EQVIB	VEFHAPFVT	RT	XWK	PWEWIWRD	XM	NTJCTYZL	OAS	XYQ

R	T	S	U	P	R	E	M
BATTALIONS	OF	ARTILLERY	MARCHING	NORTH	AT	SEVEN	OCLOCK
ARVVRKFONT	BH	SFJDUUXFP	OUVIGJPF	ULBFZ	RV	DKUKW	ROHROZ

h. In case the plain component is the reversed normal sequence, the procedure is no different from the foregoing, except that in the completion diagram the reversed sequence is employed after the cipher letters have been converted into their plain-component equivalents.

i. No doubt the student realizes from his previous work that once the primary mixed component has been recovered the latter becomes a *known* sequence and that the solution of subsequent messages employing the same set of derived alphabets, even though the keys to individual messages are different, then becomes a simple matter.

11. Solution when other types of alphabets are employed.—*a.* The foregoing examples involve the use either of standard cipher alphabets or of mixed cipher alphabets produced by the sliding of a mixed component against the normal sequence. There is, however, nothing about the general cryptographic scheme which prevents the use of other types of derived, interrelated, or secondary mixed alphabets. Cipher alphabets produced by the sliding of a mixed component against itself (either direct or reversed) or by the sliding of two different mixed components are very commonly encountered in these cases.

b. The solution of such cases involves only slight modifications in procedure, namely, those connected with the reconstruction of the primary components. The student should be in a position to employ to good advantage and without difficulty what he has learned about the principles of indirect symmetry of position in the solution of cases of the kind described.

c. The solution of a message prepared with mixed alphabets derived as indicated in subparagraph *b*, may be a difficult matter, depending upon the length of the message in question. It might, of course, be almost impossible if the message is short and there is no background for the application of the probable-word method. But if the message is quite long, or, what is more probable with respect to military communications, should the system be used for regular traffic, so that there are available for study several messages enciphered by the same set of alphabets, then the problem becomes much easier. In addition to the usual steps in solution by the probable-word method, guided by a search for and identification of idiomorphs, there is the help that can be obtained from the use of the phenomena of *isomorphism,* a study of which forms the subject of discussion in the next paragraph.

12. Isomorphism and its importance in cryptanalytics.—*a.* The term idiomorphism is familiar to the student. It designates the phenomena arising from the presence and positions of repeated letters in plain-text words, as a result of which such words may be classified according to their *compositions*, "*patterns*," or *formulae*. The term *isomorphism* (from the Greek "isos" meaning "equal" and "morphe" meaning "form") designates the phenomena arising from the existence of two or more idiomorphs with identical formulae. Two or more sequences which possess identical formulae are said to be *isomorphic.*

b. Isomorphism may exist in plain text or in cipher text. For example, the three words WARRANT, LETTERS, and MISSION are isomorphic. If enciphered monoalphabetically, their cipher equivalents would also be isomorphic. In general, isomorphism is a phenomenon of monoalphabeticity (either plain or cipher); but there are instances wherein it is latent and can be made patent in polyalphabetic cryptograms.

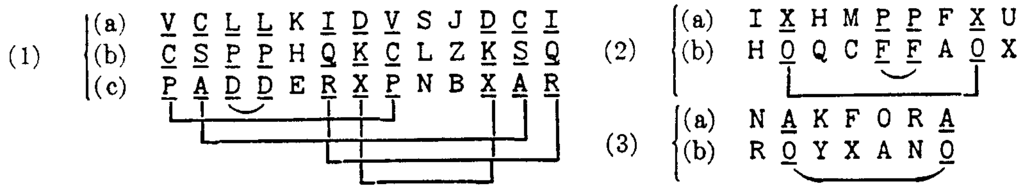
c. In practical cryptanalysis the phenomena of isomorphism afford a constantly astonishing source of clues and aids in solution. The alert cryptanalyst is always on the lookout for situations in which he can take advantage of these phenomena, for they are among the most interesting and most important in cryptanalytics.

13. Illustration of the use of isomorphism.—*a.* Let us consider the case discussed under paragraph 10, wherein a message was enciphered with a set of mixed cipher alphabets derived from sliding the key word-mixed primary component HYDRAULIC . . . XZ against the normal sequence. Suppose the message to be as follows (for simplicity, original word lengths are retained):

CRYPTOGRAM

VCLLKIDVSJDCI ORKD CFSTV IXHMPPFXU EVZZ
FK NAKFORA DKOMP ISE CSPPHQKCLZKSQ LPRO
JZWBCX HOQCFFAOX ROYXANO EMDMZMTS
TZFVUEAORSL AU PADDERXPNBXAR IGHFX JXI

b. (1) Only a few minutes inspection discloses the following three sets of isomorphs:



```
    (a) V C L L K I D V S J D C I        (2) (a) I X H M P P F X U
(1) (b) C S P P H Q K C L Z K S Q            (b) H O Q C F F A O X
    (c) P A D D E R X P N B X A R
                                         (3) (a) N A K F O R A
                                             (b) R O Y X A N O
```

(2) Without stopping to refer to word-pattern lists in an attempt to identify the very striking idiomorphs of the first set, let the student proceed to build up partial sequences of equivalents, as though he were dealing with a case of indirect symmetry of position. Thus:[2]

From isomorphs (1) (a) and (1) (b):

V⇔C; C⇔S; L⇔P; K⇔H; I⇔Q; D⇔K; S⇔L; J⇔Z;

from which the following partial sequences are constructed:

(a) VCSLP (b) DKH (c) IQ (d) JZ

From isomorphs (1) (b) and (1) (c):

C⇔P; S⇔A; P⇔D; H⇔E; Q⇔R; K⇔X; L⇔N; Z⇔B;

from which the following partial sequences are constructed:

(e) CPD (f) SA (g) HE (h) QR (i) KX (j) LN (k) ZB

From isomorphs (1) (a) and (1) (c):

V⇔P; C⇔A; L⇔D; K⇔E; I⇔R; D⇔X; S⇔N; J⇔B;

from which the following partial sequences are constructed:

(l) LDX (m) VP (n) CA (o) KE (p) IR (q) SN (r) JB

Noting that the data from the three isomorphs of this set may be combined (VCSLP and CPD make VCSLP . . D; the latter and LDX make VCSLP . . D . . . X), the following sequences are established:

	1	2	3	4	5	6	7	8	9	10	11	12	13
(1)	V	C	S	L	P	A	N	D	K	H	.	X	E

	1	2	3	4	5
(2)	I	Q	.	.	R

	1	2	3	4	5
(3)	J	Z	.	.	B

c. (1) The fact that the longest of these chains consists of exactly 13 letters and that no additions can be made from the other two cases of isomorphism, leads to the assumption that a "half-chain" is here disclosed and that the latter represents a decimation of the original primary component at an even interval. Noting the placement of the letters $\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ V & . & S & . & P & . & N & . & K \end{array}$, which gives the sequence the appearance of being the latter half of a keyword-mixed sequence running in the reversed direction, let the half-chain be reversed and extended to 26 places, as follows:

---

[2] The symbol ⇔ is to be read "is equivalent to."

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
E		.		K		N		P		S		V		X		H		D		A		L		C	

(2) The data from the two partial chains (JZ. .B and IQ. .R) may now be used, and the letters inserted into their proper positions. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
E	.	.	J	K	.	N	.	P	Q	S	.	V	.	X	Z	H	.	D	R	A	.	L	I	C	B

(3) The sequence H . D R A . L I C soon suggests HYDRAULIC as the key word. When the mixed sequence is then developed in full, complete corroboration will be found from the data of isomorphs 2 (a) (b) and 3 (a) (b). Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z

(4) From idiomorphs (2) (a) and (2) (b), the interval between H and I is 7; it is the same for O and X, Q and H, C and M, etc. From idiomorphs (3) (a) and (3) (b) the interval between R and N is 13; it is the same for O and A, Y and K, etc.

d. The message may now be solved quite readily, by the usual process of converting the cipher-text letters into their plain-component equivalents and then completing the plain component sequences. The solution is as follows:

[Key: STRIKE WHILE THE IRON IS . . . (HOT?)]

```
            S                 T         R              I             K
C O M M U N I C A T I O N  W I T H  F I R S T  A R T I L L E R Y  W I L L
V C L L K I D V S J D C I  O R K D  C F S T V  I X H M P P F X U  E V Z Z

 E         W            H        I                L                 E
B E  T H R O U G H  C O R P S  A N D  C O M M U N I C A T I O N  W I T H
F K  N A K F O R A  D K O M P  I S E  C S P P H Q K C L Z K S Q  L P R O

     T                H               E               I
S E C O N D  A R T I L L E R Y  T H R O U G H  D I V I S I O N
J Z W B C X  H O Q C F F A O X  R O Y X A N O  E M D M Z M T S

          R             O               N                   I       S
S W I T C H B O A R D  N O  C O M M U N I C A T I O N  A F T E R  T E N
T Z F V U E A O R S L  A U  P A D D E R X P N B X A R  I G H F X  J X I
```

e. (1) In the foregoing illustration the steps are particularly simple because of the following circumstances:

(a) The actual word lengths are shown.

(b) The words are enciphered monoalphabetically by different alphabets belonging to a set of secondary alphabets.

(c) Repetitions of plain-text words, enciphered by different alphabets, produce isomorphs and the lengths of the isomorphs are definitely known as a result of circumstance (a).

(2) Of these facts, the last is of most interest in the present connection. But what if the actual word lengths are not shown; that is, what if the text to be solved is intercepted in the usual 5-letter-group form?

462861 O - 42 - 2

SECTION IV

SOLUTION OF SYSTEMS USING CONSTANT-LENGTH KEYING UNITS TO ENCIPHER VARIABLE-LENGTH PLAIN-TEXT GROUPINGS, III

Paragraph
General remarks........ 14
Word separators........ 15
Variations and concluding remarks on foregoing systems........ 16

14. General remarks.—*a.* The cases described thus far are particularly easy to solve because the cryptanalyst has before him the messages in their true or original word lengths. But in military cryptography this is seldom or never the case. The problem is therefore made somewhat more difficult by reason of the fact that there is nothing to indicate definitely the limits of encipherment by successive keyletters. However, the solution merely necessitates more experimentation in this case than in the preceding. The cryptanalyst must take careful note of repetitions which may serve to "block out" or delimit words, and hope that when this is done he will be able to find and identify certain sequences having familiar idiomorphic features or patterns, such as those noted above. If there is plenty of text, repetitions will be sufficient in number to permit of employing this entering wedge.

b. Of course, if any sort of stereotypic phraseology is employed, especially at the beginnings or endings of the messages, the matter of assuming values for sequences of cipher letters is easy, and affords a quick solution. For example, suppose that as a result of previous work it has been found that many messages begin with the expression REFERRING TO YOUR NUMBER . . . . Having several messages for study, the selection of one which begins with such a common idiomorphism as that given by the word REFERRING is a relatively simple matter; and having found the word REFERRING, if with a fair degree of certainty one can add the words TO YOUR NUMBER, the solution is probably well under way.

c. (1) Take the case discussed in paragraph 13, but assume that word lengths are no longer indicated because the message is transmitted in the usual 5-letter groups. The process of ascertaining the exact length of sequences which are isomorphic, or, as the process is briefly termed, "blocking out isomorphs" becomes a more difficult matter and must often rest upon rather tenuous threads of reasoning. For example, take the illustrative message just dealt with and let it be assumed that it was arranged in 5-letter groups.

```
VCLLK  IDVSJ  DCIOR  KDCFS  TVIXH  MPPFX
UEVZZ  FKNAK  FORAD  KOMPI  SECSP  PHQKC
LZKSQ  LPROJ  ZWBCX  HOQCF  FAOXR  OYXAN
OEMDM  ZMTST  ZFVUE  AORSL  AUPAD  DERXP
NBXAR  IGHFX  JXI
```

(2) The detection of isomorphisms now becomes a more difficult matter. There is no special trouble in picking out the following three isomorphic sequences:

```
(1) V C L L K I D V S J D C I
(2) C S P P H Q K C L Z K S Q
(3) P A D D E R X P N B X A R
```

since the first one happens to be at the beginning of the message and its left-hand boundary, or "head," is marked by (or rather, coincides with) the beginning of the message. By a fortunate circumstance, the right-hand boundary, or "tail," can be fixed just as accurately. That the repetition extends as far as indicated above is certain for we have a check on the last column I, Q, R. If an additional column were added, the letters would be O, L, I. Since the second letter has previously appeared while the first and third have not, a contradiction results and the new column may not be included.

If, however, none of the three letters O, L, I had previously appeared, so that there could be no means of getting a check on their correctness, it would not be possible to block out or ascertain the extent of the isomorphism in such a case. All that could be said would be that it seems to include the first 13 letters, *but it might continue further.*

d. (1) However, the difficulty or even the impossibility of blocking out the isomorphs *to their full extent* is not usually a serious matter. After all, the cryptanalyst uses the phenomenon not to identify words but to obtain cryptanalytic data for reconstructing cipher alphabets. For example, how many data are lost when the illustrative message of subparagraph 13*a* is rewritten in 5-letter groups as in subparagraph 14*c* (1)? Suppose the latter form of message be studied for isomorphs:

```
VCLLK  IDVSJ  DCIOR  KDCFS  TVIXH  MPPFX  UEVZZ
FKNAK  FORAD  KOMPI  SECSP  PHQKC  LZKSQ  LPROJ
ZWBCX  HOQCF  FAOXR  OYXAN  OEMDM  ZMTST  ZFVUE
AORSL  AUPAD  DERXP  NBXAR  IGHFX  JXI
```

(2) If the underscored sequences are compared with those in the message in subparagraph 13*a*, it will be found that only a relatively small amount of information has been lost. Certainly not enough to cause any difficulty have been lost in this case, for all the data necessary for the reconstruction of the mixed cipher component came from the first set of isomorphs, and the latter are identical in length in both cases. Only the head and tail letters of the second pair of isomorphic sequences are not included in the underscored sequences in the 5-letter version of the message. The third pair of isomorphic sequences shown in paragraph 13*b* does not appear in the 5-letter version since there is only one repeated letter in this case. In long messages or when there are many short messages, a study of isomorphism will disclose a sufficient number of partial isomorphs to give data usually sufficient for purposes of alphabet reconstruction.

e. It should be noted that there is nothing about the phenomenon of isomorphism which restricts its use to cases in which the cipher alphabets are secondary alphabets resulting from the sliding of a mixed component against the normal. It can be useful in *all* cases of interrelated secondary alphabets no matter what the basis of their derivation may be.

f. In subsequent studies the important role which the phenomenon of isomorphism plays in cryptanalytics will become more apparent. When the traffic is stereotypic in character, even to a slight degree, so that isomorphism may extend over several words or phrases, the phenomenon becomes of highest importance to the cryptanalyst and an extremely valuable tool in his hands.

15. Word separators.—*a.* One of the practical difficulties in employing systems in which the keying process shifts according to word lengths is that in handling such a message the decryptographing clerk is often not certain exactly when the termination of a word has been reached, and thus time is lost by him. For instance, while decryptographing a word such as INFORM the clerk would not know whether he now has the complete word and should shift to the next key letter or not: The word might be INFORMS, INFORMED, INFORMING, INFORMAL, INFOR-

MATION, etc. The past tense of verbs, the plural of nouns, and terminations of various sorts capable of being added to word roots would give rise to difficulties, and the latter would be especially troublesome if the messages contained a few telegraphic errors. Consequently, a scheme which is often adopted to circumvent this source of trouble is to indicate the end of a word by an infrequent letter such as Q or X, and enciphering the letter. In such usage these letters are called *word separators*.

b. When word separators are employed and this fact is once discovered, their presence is of as much aid to the cryptanalyst in his solution as it is to the clerks who are to decryptograph the messages. Sometimes the presence of these word separators, even when enciphered, aids or makes possible the blocking out of isomorphs.

16. Variations and concluding remarks on foregoing systems.—*a*. The systems thus far described are all based upon word-length encipherment using different cipher alphabets. Words are markedly irregular in regard to this feature of their construction, and thus aperiodicity is imparted to such cryptograms. But variations in the method, aimed at making the latter somewhat more secure, are possible. Some of these variations will now be discussed.

b. Instead of enciphering according to natural word lengths, the irregular groupings of the text may be regulated by other agreements. For example, suppose that the numerical value (in the normal sequence) of each key letter be used to control the number of letters enciphered by the successive cipher alphabets. Depending then upon the composition of the key word or key phrase, there would be a varying number of letters enciphered in each alphabet. If the key word were PREPARE, for instance, then the first cipher alphabet would be used for 16 (P=16) letters, the second cipher alphabet, for 18 (=R) letters, and so on. Monoalphabetic encipherment would therefore allow plenty of opportunity for tell-tale word patterns to manifest themselves in the cipher text. Once an entering wedge is found in this manner, solution would be achieved rather rapidly. Of course, all types of cipher alphabets may be employed in this and the somewhat similar schemes described.

c. If the key is short, and the message is long, periodicity will be manifested in the cryptogram, so that it would be possible to ascertain the length of the basic cycle (in this case the length of the key) despite the irregular groupings in encipherment. The determination of the length of the cycle might, however, present difficulties in some cases, since the basic or fundamental period would not be clearly evident because of the presence of repetitions which are not periodic in their origin. For example, suppose the word PREPARE were used as a key, each key letter being employed to encipher a number of letters corresponding to its numerical value in the normal sequence. It is clear that the length of the basic period, in terms of letters, would here be the sum of the numerical values of P (=16)+R (=18) + E (=5), and so on, totalling 79 letters. But because the key itself contains repeated letters and because encipherment by each key letter is monoalphabetic there would be plenty of cases in which the first letter P would encipher the same or part of the same word as the second letter P, producing repetitions in the cryptogram. The same would be true as regards encipherments by the two R's and the two E's in this key word. Consequently, the basic period of 79 would be distorted or masked by aperiodic repetitions, the intervals between which would not be a function of, nor bear any relation to, the length of the key. The student will encounter more cases of this kind, in which a fundamental periodicity is masked or obscured by the presence of cipher-text repetitions not attributable to the fundamental cycle. The experienced cryptanalyst is on the lookout for phenomena of this type, when he finds in a polyalphabetic cipher plenty of repetitions but with no factorable constancy which leads to the disclosure of a short period. He may conclude, then, either that the cryptogram involves several primary periods which interact to produce a long resultant period, or that it involves a fairly long fundamental cycle within which repetitions of a nonperiodic origin are present and obscure the phenomena manifested by repetitions of a periodic origin.

d. (1) A logical extension of the principle of polyalphabetic encipherment of variable-length plain-text groupings is the case in which these plain-text groupings rarely exceed 4 letters, so that a given cipher alphabet is in play for only a very short time, thus breaking up what might otherwise appear as fairly long repetitions in the cipher text. For example, suppose the letters of the alphabet, arranged in their normal-frequency order, were set off into four groups, as follows:

E T R I N	O A S D L C	H F U P M Y G	W V B X K Q J Z
Group 1	Group 2	Group 3	Group 4

(2) Suppose that a letter in group 1 means that one letter will be enciphered; a letter in group 2, that two letters will be enciphered; and so on. Suppose, next, that a rather lengthy phrase were used as a key, for example, PREPARED UNDER THE DIRECTION OF THE CHIEF SIGNAL OFFICER FOR USE WITH ARMY EXTENSION COURSES. Suppose, finally, that each letter of the key were used not only to select the particular cipher alphabet to be used, but also to control the number of letters to be enciphered by the selected alphabet, according to the scheme outlined above. Such an enciphering scheme, using the HYDRAULIC...XZ primary cipher component sliding against the normal plain component, would yield the following groupings:

Grouping	3	1	1	3	2	1	1	2	3	1	2	1	1	1	3	1	2
Key	P	R	E	P	A	R	E	D	U	N	D	E	R	T	H	E	D
Plain	FIR	S	T	DIV	IS	I	O	NW	ILL	A	DV	A	N	C	EAT	F	IV
Cipher	WHB	T	R	THJ	GV	F	X	MX	JNN	N	UW	E	N	W	AHQ	M	EW

Grouping	1	1	1	2	1	1	2	1	2	3	1	3	1	2	3	1	
Key	I	R	E	C	T	I	O	N	O	F	T	H	E	C	H	I	...
Plain	E	F	I	FT	E	E	NA	M	AS	SEC	O	NDD	I	VI	SIO	N	...
Cipher	F	C	P	JY	Z	F	AO	D	OB	RMJ	B	JRR	P	RN	PCK	S	...

(3) Here it will be seen that any tendency for the formation of lengthy repetitions would be counteracted by the short groupings and quick shifting of alphabets. The first time the word DIVISION occurs it is enciphered as THJGVFXM; the second time it occurs it is enciphered as RPRNPCKS. Before DIVISION can be twice enciphered by *exactly* the same sequence of key letters, an interval of at least 140 letters must intervene between the two occurrences of the word (the sum of the values of the letters of the key phrase=140); and then the chances that the key letter P would begin the encipherment of DIV are but one in three. Only one of these three possible encipherments will yield *exactly* the same sequence of cipher equivalents the second time as was obtained the first time. For example, if the text were such as to place two occurrences of the word DIVISION in the positions shown below, their encipherments would be as follows:

3	1	1	3	2	1	1	2	3	1
P	R	E	P	A	R	E	D	U	N
FIR	S	T	DIV	IS	I	O	N	...	.
...	.	.	THJ	GV	F	X	M	...	.

3	1	1	3	2	1	1	2	3	1
P	R	E	P	A	R	E	D	U	N
...	.	.	.DI	VI	S	I	ON	...	.
...	.	.	.TH	ZG	T	P	NM	...	.

Although the word DIVISION, on its second appearance, begins but one letter beyond the place where it begins on its first appearance, the cipher equivalents now agree only in the first two letters, the fourth, and the last letters. Thus:

```
      D I V I S I O N
(1)   T H J G V F X M
(2)   T H Z G T P N M
```

e. Attention is directed to the characteristics of the foregoing two encipherments of the same word. When they are superimposed, the first two cipher equivalents are the same in the two encipherments; then there is a single interval where the cipher equivalents are different; the next cipher equivalent is the same; then follow three intervals with dissimilar cipher equivalents; finally, the last cipher equivalent is the same in both cases. The repetitions here extend only to one or two letters; longer repetitions can occur only exceptionally. The two encipherments yield only occasional *coincidences*, that is, places where the cipher letters are identical; moreover, the *distribution* of the coincidences is quite irregular and of an intermittent character.

f. This phenomenon of *intermittent coincidences*, involving coincidences of single letters, pairs of letters, or short sequences (rarely ever exceeding pentagraphs) is one of the characteristics of this general class of polyalphabetic substitution, wherein the cryptograms commonly manifest what appears to be a disturbed or distorted periodicity.

g. From a technical standpoint, the cryptographic principle upon which the foregoing system is based has much merit, but for practical usage it is entirely too slow and too subject to error. However, if the encipherment were mechanized by machinery, and if the enciphering key were quite lengthy, such a system and mechanism becomes of practical importance. Cipher machines for accomplishing this type of substitution will be treated in a subsequent text.

SECTION V

SOLUTION OF SYSTEMS USING VARIABLE-LENGTH KEYING UNITS TO ENCIPHER CONSTANT-LENGTH PLAIN-TEXT GROUPINGS

Paragraph
Variable-length groupings of the keying sequence 17
Methods of interrupting a cyclic keying sequence 18
Interruptor is a plain-text letter 19
Solution by superimposition 20
Interruptor is a cipher-text letter 21
Concluding remarks 22

17. Variable-length groupings of the keying sequence.—The preceding cases deal with simple methods of eliminating or avoiding periodicity by enciphering variable-length groupings of the plain text, using constant-length keying units. In paragraph 2*a*, however, it was pointed out that periodicity can also be suppressed by applying variable-length key groupings to constant-length plain-text groups. One such method consists in *irregularly interrupting* the keying sequence, if the latter is of a limited or fixed length, and recommencing it (from its initial point) after such interruption, so that the keying sequence becomes equivalent to a series of keys of different lengths. Thus, the key phrase BUSINESS MACHINES may be expanded to a series of irregular-length keying sequences, such as BUSI/BUSINE/BU/BUSINESSM/BUSINESSMAC, etc. Various schemes or prearrangements for indicating or determining the interruptions may be adopted. Three methods will be mentioned in the next paragraph.

18. Methods of interrupting a cyclic keying sequence.—*a*. There are many methods of interrupting a keying sequence which is basically cyclic, and which therefore would give rise to periodicity if not interfered with in some way. These methods may, however, be classified into three categories as regards what happens after the interruption occurs:

(1) The keying sequence merely stops and begins again at the initial point of the cycle.

(2) One or more of the elements in the keying sequence may be omitted from time to time irregularly.

(3) The keying sequence irregularly alternates in its direction of progression, with or without omission of some of its elements.

b. These methods may, for clarity, be represented graphically as follows. Suppose the key consists of a cyclic sequence of 10 elements represented symbolically by the series of numbers 1, 2, 3, . . ., 10. Using an asterisk to indicate an interruption, the following may then represent the relation between the letter number of the message and the element number of the keying sequences in the three types mentioned above:

(1)	Letter No.	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
	Key element No.	1-2-3-4-*-1-2-3-4-5- 6-*- 1- 2- 3-*- 1- 2- 3- 4- 5- 6- 7-*
	Letter No.	21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
	Key element No.	1- 2- 3- 4- 5- 6- 7- 8- 9-10-*- 1- 2- 3-*- 1- 2- etc.
(2)	Letter No.	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
	Key element No.	1-2-3-*-7-8-9-10-1-2-*- 4- 5- 6-*- 3- 4- 5- 6- 7- 8- 9-10-
	Letter No.	21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
	Key element No.	1-*- 8- 9-10- 1- 2-*- 5- 6- 7-*- 9-10- 1-*- 5- 6- 7- etc.

(3)
Letter No.................. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Key element No.......... 1-2-3-4-5-*-4-3-*-4-5- 6- 7- 8- 9-10- 1-*-10- 9- 8- 7-*- 8
Letter No.................. 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Key element No.......... 9-10- 1- 2- 3-*-2- 1-10- 9- 8-*- 9-10- 1- 2- 3 etc.

As regards the third method, which involves only an alternation in the direction of progression of the keying sequence, if there were no interruptions in the key it would mean merely that a 10-element keying sequence, for example, could be treated as though it were an 18-element sequence and the matter could then be handled as though it were a special form of the second method. But if the principles of the second and third method are combined in one system, the matter may become quite complex.

c. If one *knows* when the interruptions take place in each cycle, then successive *sections* of the basic keying cycle in the three cases may be superimposed. Thus:

METHOD (1)

Keying element No..	1	2	3	4	5	6	7	8	9	10
Letter No..................	1	2	3	4 |						
Letter No..................	5	6	7	8	9	10 |				
Letter No..................	11	12	13 |							
Letter No..................	14	15	16	17	18	19	20 |			
Letter No..................	21	22	23	24	25	26	27	28	29	30 |
Letter No..................	31	32	33 |							
Letter No..................	34	35	etc.							

METHOD (2)

Keying element No..	1	2	3	4	5	6	7	8	9	10
Letter No..................	1	2	3 |	—	—	—	4	5	6	7
Letter No..................	8	9 |	—	10	11	12 |				
Letter No..................	—	—	13	14	15	16	17	18	19	20
Letter No..................	21 |	—	—	—	—	—	—	22	23	24
Letter No..................	25	26 |	—	—	27	28	29 |	—	30	31
Letter No..................	32 |	—	—	—	33	34	35	etc.		

METHOD (3)

Keying element No..	1	2	3	4	5	6	7	8	9	10
Letter No..................	1	2	3	4	5 |	—	—	—	—	—
Letter No..................	—	— |	7	6	—	—	—	—	—	—
Letter No..................	—	—	—	8	9	10	11	12	13	14
Letter No..................	15 |	—	—	—	—	— |	19	18	17	16
Letter No..................	23	24	25 |	—	—	—	—	20	21	22
Letter No..................	27	26	—	—	—	—	— |	30	29	28
Letter No..................	33	34	35	etc.					31	32

Obviously if one does not know when or how the interruptions take place, then the successive sections of keying elements cannot be superimposed as indicated above.

d. The interruption of the cyclic keying sequence usually takes place according to some prearranged plan, and the three basic methods of interruption will be taken up in turn, using a short mnemonic key as an example.

e. Suppose the correspondents agree that the interruption in the keying sequence will take place after the occurrence of a specified letter called an *interruptor*,[1] which may be a letter of the plain text, or one of the cipher text, as agreed upon in advance. Then, since in either case there is nothing fixed about the time the interruption will occur—it will take place at no fixed intervals—not only does the interruption become quite irregular, following no pattern, but also the method never reverts to one having periodicity. Methods of this type will now be discussed in detail.

19. Interruptor is a plain-text letter.—*a.* Suppose the correspondents agree that the interruption in the key will take place immediately after a previously agreed-upon letter, say R, occurs in the plain text. The key would then be interrupted as shown in the following example (using the mnemonic key BUSINESS MACHINES and the HYDRAULIC . . . XZ sequence):

Key	B U S I N E S S M A C H I	B U S	B U S I	B U S I N E
Plain	A M M U N I T I O N F O R	F I R	S T A R	T I L L E R
Cipher	B O L Y R P J D R O J K X	K J F	Y X S X	D J U P S Y

Key	B U S I N E S S M A C H I N E S B U	B U S I N E S S M A C H I
Plain	Y W I L L B E L O A D E D A F T E R	A M M U N I T I O N F O R
Cipher	I Y D P Y F X U R A F A E N M J J V	B O L Y R P J D R O J K X

Key	B U S I	B U S	B U S I N E	B U S I N
Plain	T H I R	D A R	T I L L E R	Y . . . .
Cipher	D G D X	G U F	D J U P S Y	I . . . .

CRYPTOGRAM

```
BOLYR  PJDRO  JKXKJ  FYXSX  DJUPS  YIYDP
YFXUR  AFAEN  MJJVB  OLYRP  JDROJ  KXDGD
XGUFD  JUPSY  IXXXX
```

b. Instead of employing an ordinary plain-text letter as the interruptor, one might reserve the letter J for this purpose (and use the letter I whenever this letter appears as part of a plain-text word). This is a quite simple variation of the basic method. The letter J acts merely as though it were a plain-text letter, except that in this case it also serves as the interruptor. The interruptor is then inserted *at random*, at the whim of the enciphering clerk. Thus:

Key	BUSINESSMAC	BUSINESSM	BUSINESSMACHINESBUSIN
Plain	TROOPSWILLJ	BEHALTEDJ	ATROADIUNCTIONFIVESIX

c. It is obvious that repetitions would be plentiful in cryptograms of this construction, regardless of whether a letter of high, medium, or low frequency is selected as the signal for key interruption. If a letter of high frequency is chosen, repetitions will occur quite often, not only because that letter will certainly be a part of many common words, but also because it will be followed by words that are frequently repeated; and since the key starts again with each such interruption, these frequently repeated words will be enciphered by the same sequence of alphabets. This is the case in the first of the two foregoing examples. It is clear, for instance, that every time the word ARTILLERY appears in the cryptogram the cipher equivalents of TILLERY must be the same. If the interruptor letter were A_p instead of R_p, the repetition would include the cipher equivalents of RTILLERY; if it were T_p, ILLERY, and so on. On the other hand, if a letter of low frequency were selected as the interruptor letter, then the encipherment would tend to approximate that of normal repeating-key substitution, and repetitions would be plentiful on that basis alone.

---

[1] Also called at times an "influence" letter because it influences or modifies normal procedure. In some cases no influence or interruptor letter is used, the interruption or break in the keying sequence occurring after a previously-agreed-upon number of letters has been enciphered.

d. Of course, the lengths of the intervals between the repetitions, in any of the foregoing cases, would be irregular, so that periodicity would not be manifested. The student may inquire, therefore, how one would proceed to solve such messages, for it is obvious that an attempt to allocate the letters of a single message into separate monoalphabetic distributions cannot be successful unless the exact locations of the interruptions are known—and they do not become known to the cryptanalyst until he has solved the message, or at least a part of it. Thus it would appear as though the would-be solver is here confronted with a more or less insoluble dilemma. This sort of reasoning, however, makes more of an appeal to the novice in cryptography than to the experienced cryptanalyst, who specializes in methods of solving cryptographic dilemmas.

e. (1) The problem here will be attacked upon the usual two hypotheses, and the easier one will be discussed first. Suppose the system has been in use for some time, that an original solution has been reached by means to be discussed under the second hypothesis, and that the cipher alphabets are known. There remains unknown only the specific key to messages. Examining whatever repetitions are found, an attack is made on the basis of searching for a probable word. Thus, taking the illustrative message in subparagraph *a*, suppose the presence of the word ARTILLERY is suspected. Attempts are made to locate this word, basing the search upon the construction of an intelligible key. Beginning with the very first letter of the message, the word ARTILLERY is juxtaposed against the cipher text, and the key letters ascertained, using the known alphabets, which we will assume in this case are based upon the HYDRAULIC . . . XZ sequence sliding against the normal. Thus:

Cipher	B	O	L	Y	R	P	J	D	R
Plain	A	R	T	I	L	L	E	R	Y
"Key"	B	H	J	Q	P	I	B	F	U

(2) Since this "key" is certainly not intelligible text, the assumed word is moved one letter to the right and the test repeated, and so on until the following place in the test is reached:

Cipher	S	X	D	J	U	P	S	Y	I
Plain	A	R	T	I	L	L	E	R	Y
Key	S	I	B	U	S	I	N	E	B

(3) The sequence BUSINE suggests BUSINESS; moreover, it is noted that the key is interrupted both times by the letter R_p. Now the key may be applied to the beginning of the message, to see if the whole key or only a portion of it has been recovered. Thus:

Key	B	U	S	I	N	E	S	S	B	U	S
Cipher	B	O	L	Y	R	P	J	D	R	O	J
Plain	A	M	M	U	N	I	T	I	U	M	T

(4) It is obvious that BUSINESS is only a part of the key. But the deciphered sequence certainly seems to be the word AMMUNITION. When this is tried, the key is extended to BUSINESS MA . . . . Enough has been shown to clarify the procedure.

f. The foregoing solution is predicated upon the hypothesis that the cipher alphabets are known. But what if this is not the case? What of the steps necessary to arrive at the *first* solution, before even the presence of an interruptor is suspected? The answer to this question leads to the presentation of a method of attack which is one of the most important and powerful means the cryptanalyst has at his command for unraveling many knotty problems. It is called *solution by superimposition*, and warrants detailed treatment.

20. Solution by superimposition.— *a.* *Basic principles.* — (1) In solving an ordinary repeating-key cipher the first step, that of ascertaining the length of the period, is of no significance in itself. It merely paves the way for and makes possible the second step, which consists in allocating the letters of the cryptogram into individual monoalphabetic distributions. The third step then consists in solving these distributions. Usually, the text of the message is transcribed into its periods and is written out in successive lines corresponding in length with that of the period. The diagram then consists of a series of columns of letters and the letters in each column belong to the same monoalphabet. Another way of looking at the matter is to conceive of the text as having thus been transcribed into *superimposed periods:* in such case the letters in each column have undergone the same kind of treatment by the same elements (plain and cipher components of the cipher alphabet).

(2) Suppose, however, that the repetitive key is very long and that the message is short, so that there are only a very few cycles in the text. Then the solution of the message becomes difficult, if not impossible, because there is not a sufficient number of superimposable periods to yield monoalphabetic distributions which can be solved by frequency principles. But suppose also that there are many short cryptograms all enciphered by the same key. Then it is clear that if these messages are superimposed:

(a) The letters in the respective columns will all belong to individual alphabets; and

(b) If there is a sufficient number of such superimposable messages (say 25–30, for English), then the frequency distributions applicable to the successive *columns* of text can be solved—*without knowing the length of the key.* In other words, any difficulties that may have arisen on account of failure or inability to ascertain the length of the period have been circumvented. The second step in normal solution is thus "by-passed."

(3) Furthermore, and this is a very important point, in case an extremely long key is employed and a series of messages beginning at different initial points are enciphered by such a key, this method of solution by superimposition can be employed, provided the messages can be superimposed correctly, that is, so that the letters which fall in one column really belong to one cipher alphabet. Just how this can be done will be demonstrated in subsequent paragraphs, but a clue has already been given in paragraph 18*c*. At this point, however, a simple illustration of the method will be given, using the substitution system discussed in paragraph 19.

b. *Example.*—(1) A set of 35 messages has been intercepted on the same day. Presumably they are all in the same key, and the presence of repetitions between messages corroborates this assumption. But the intervals between repetitions within the same message do not show any common factor and the messages appear to be aperiodic in nature. The probable-word method has been applied, using standard alphabets, with no success. The messages are then superimposed (Fig. 5); the frequency distributions for the first 10 columns are as shown in Figure 6.

1 Z C T P Z W Z P E P Z Q X
2 W T E Q M X Z S Y S P R C
3 T C R W C X T B H H
4 E F K C S Z R I H A
5 Y A N C I H Z N U W
6 V Z I E T I R R G X
7 H C Q I C K G U O N
8 Z C F C L X R K Q W
9 H W W P T E W C I M J S
10 E P D O Z C L I K S J
11 W T S S Q Z P Z I E T
12 Z C G G Y F C S B G
13 C W Z A O O E M H W T P
14 C I Y G I F B D T V X
15 E A Q D R D N S R C A P D T
16 Y F W C Q Q B Z C W C
17 W T E Z Q S K U H C
18 Z C V X Q Z K Z Y D W L K
19 A F E O J T D T I T
20 K P V F Q W P K T E V
21 Z A B G R T X P U Q X
22 Y H E O C U H M D T
23 C L C P Z I K O T H
24 A F L W W Z Q M D T
25 Z C W A P M B S A W L
26 H F L M H R Z N A P E C E
27 C L Z G E M K Z T O
28 T P Y F K O T I Z U H
29 Z C C P S N E O P H D Y L
30 C I Y G I F T S Y T L E
31 Y T S V W V D G H P G U Z
32 N O C A I F B J B L G H Y
33 Z X X F L F E G J L
34 Z C T M M B Z J O O
35 H C Q I W S Y S B P H C Z V

FIGURE 5.

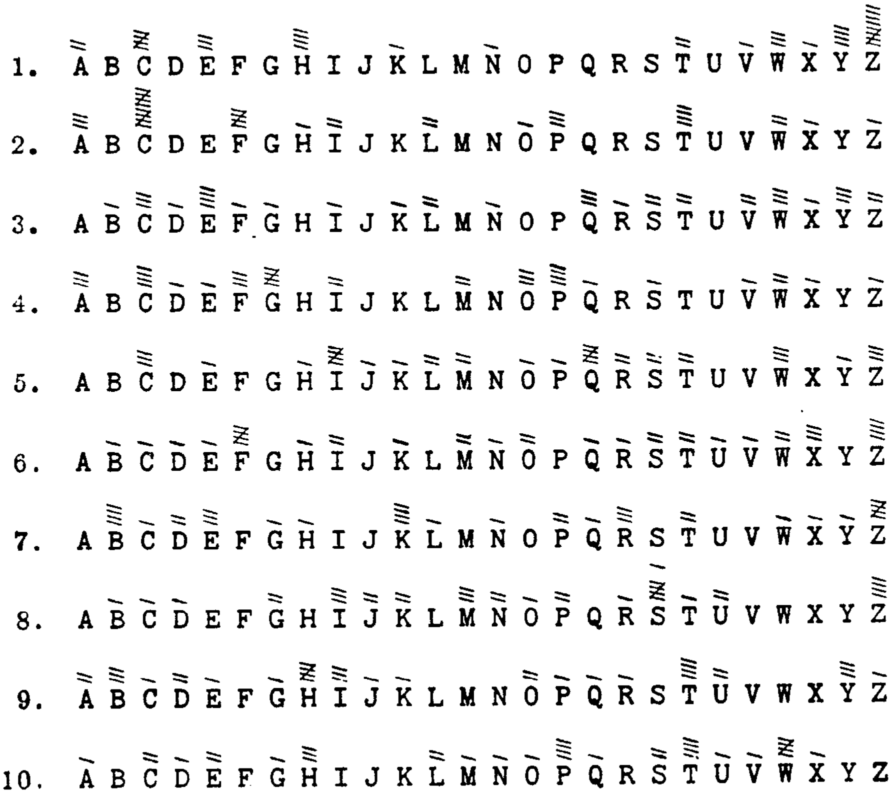

1. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
2. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
3. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
4. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
5. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
6. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
7. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
8. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
9. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
10. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

FIGURE 6.

(2) The 1st and 2d distributions are certainly monoalphabetic. There are very marked crests and troughs, and the number of blanks (14) is more than satisfactory in both cases. (Let the student at this point refer to Par. 14 and Chart 5 of Military Cryptanalysis, Part I.) But the 3d, 4th, and remaining distributions appear no longer to be monoalphabetic. Note particularly the distribution for the 6th column. From this fact the conclusion is drawn that some disturbance in periodicity has been introduced in the cryptograms. In other words, although they all start out with the same alphabet, some sort of interruption takes place so as to suppress periodicity.

(3) However, a start on solution may be made by attacking the first two distributions, frequency studies being aided by considerations based upon probable words. In this case, since the text comprises only the beginnings of messages, assumptions for probable words are more easily made than when words are sought in the interiors of messages. Such common introductory words as REQUEST, REFER, ENEMY, WHAT, WHEN, IN, SEND, etc., are good ones to assume. Furthermore, high-frequency digraphs used as the initial digraphs of common words will, of course, manifest themselves in the first two columns. The greatest aid in this process is, as usual, a familiarity with the "word habits" of the enemy.

(4) Let the student try to solve the messages. In so doing he will more or less quickly find the cause of the rapid falling off in monoalphabeticity as the columns progress to the right from the initial point of the messages.

21. Interruptor is a cipher-text letter.—*a.* In the preceding case a plain-text letter serves as the interruptor. But now suppose the correspondents agree that the interruption in the key will take place immediately after a previously-agreed-upon letter, say Q, occurs in the cipher text. The key would then be interrupted as shown in the following example:

```
Key.................. B U S I N E S S M A C H I N E S B U S I N E S S M|
Plain................ A M M U N I T I O N F O R F I R S T A R T I L L E|
Cipher............... B O L Y R P J D R O J K X T P F Y X S X B P U U Q|

Key.................. B U S I N E S S M A C H I N|B U S I N E S S M A C H|B U|
Plain................ R Y W I L L B E L O A D E D|A F T E R A M M U N I T|I O|
Cipher............... H R N M Y T T X H P C R F Q|B E J F I E L L B O N Q|O Q|

Key.................. B U S I N E S S M A C H|B U S I N E
Plain................ N F O R T H I R D A R T|I L L E R Y
Cipher............... V E C X B O D F P A Z Q|O N U F I C
```

CRYPTOGRAM

```
BOLYR  PJDRO  JKXTP  FYXSX  BPUUQ  HRNMY
TTXHP  CRFQB  EJFIE  LLBON  QOQVE  CXBOD
FPAZQ  ONUFI  CXXXX
```

b. In the foregoing example, there are no significant repetitions. Such as do occur comprise only digraphs, one of which is purely accidental. But the absence of significant, long repetitions is itself purely accidental; for had the interruptor letter been a letter other than Q_c, then the phrase AMMUNITION FOR, which occurs twice, might have been enciphered identically both times. If a short key is employed, repetitions may be plentiful. For example, note the following, in which S_c is the interruptor letter:

Key	B A N D S B A N D S B A N D S B A N D S B A N	B A N D S B A N D S B
Plain	F R O M F O U R F I V E T O F O U R F I F T E	E N A M B A R R A G E
Cipher	K T A K $\underline{\text{Z W X I I D}}$ A C B N $\underline{\text{Z W X I I D}}$ K W S	J O N K T B T I D H J

c. This last example gives a clue to one method of attacking this type of system. There will be repetitions within short sections, and the interval between them will sometimes permit of ascertaining the length of the key. In such short sections, the letters which intervene between the repeated sequences may be eliminated as possible interruptor letters. Thus, the letters A, C, B, and N may be eliminated, in the foregoing example, as interruptor letters. By extension of this principle to the letters intervening between other repetitions, one may more or less quickly ascertain what letter serves as the interruptor.

d. Once the interruptor letter has been found, the next step is to break up the message into "uninterrupted" sequences and then attempt a solution by superimposition. The principles explained in paragraph 20 need only be modified in minor respects. In the first place, in this case the columns of text formed by the superimposition of uninterrupted sequences will be purely monoalphabetic, whereas in the case of the example in paragraph 20, only the very first column is purely monoalphabetic, the monoalphabeticity falling off very rapidly with the 2d, 3d, . . . columns. Hence, in this case the analysis of the individual alphabets should be an easier task. But this would be counterbalanced by the fact that whereas in the former case the cryptanalyst is dealing with the initial words of messages, in this case he is dealing with interior portions of the text and has no way of knowing where a word begins. The latter remarks naturally do not apply to the case where a whole set of messages in this system, all in the same key, can be subjected to simultaneous study. In such a case the cryptanalyst would also have the initial words to work upon.

22. **Concluding remarks.**—*a.* The preceding two paragraphs both deal with the first and simplest of the three basic cases referred to under paragraph 18. The second of those cases involves considerably more work in solution for the reason that when the interruption takes place and the keying sequence recommences, the latter is not invariably the initial point of the sequence, as in the first case.

b. In the second of those cases the interruptor causes a break in the keying sequence and a recommencement at any one of the 10 keying elements. Consequently, it is impossible now merely to superimpose sections of the text by shifting them so that their initial letters fall in the same column. But a superimposition is nevertheless possible, provided the interruptions do not occur so frequently [2] that sections of only a very few letters are enciphered by sequent keyletters. In order to accomplish a proper superimposition in this case, a statistical test is essential, and for this a good many letters are required. The nature of this test will be explained in Section XI.

c. The same thing is true of the last of the three cases mentioned under paragraph 18. The solution of a case of this sort is admittedly a rather difficult matter which will be taken up in its proper place later.

d. (1) In the cases thus far studied, either the plain-text groupings were variable in length and were enciphered by a constant-length key, or the plain-text groupings were constant in length and were enciphered by a variable-length key. It is possible, however, to combine both principles and to apply a variable-length key to variable-length groupings of the plain text.

---

[2] When no interruptor or "influence letter" is used, the interruption or break in the keying sequence occurs after the encipherment of a definite number of letters. Once this number has been ascertained, solution of subsequent messages is very simple.

(2) Suppose the correspondents agree to encipher a message according to word lengths, but at irregular intervals, to add at the end of a word an interruptor letter which will serve to interrupt the key. Note the following, in which the key is BUSINESS MACHINES and the interruptor letter is X:

Key	B	U	S	B	
Plain	A M M U N I T I O N	F O R	F I R S T X	A R T I L L E R Y	etc.
Cipher	B T T R V O D O W V	E Q V	Z D F G J O	B H D O S S J H I	

CRYPTOGRAM

BTTRV ODOWV EQVZD FGJOB HDOSS JHI . . . etc.

(3) The foregoing system is only a minor modification of the simple case of ordinary word length encipherment as explained in Section II. If standard cipher alphabets are used, the spasmodic interruption and the presence of the interruptor letter would cause no difficulty whatever, since the solution can be achieved mechanically, by completing the plain-component sequence. If mixed cipher alphabets are used, and the primary components are unknown, solution may be reached by following the procedure outlined in Sections II and III, with such modifications as are suitable to the case.

e. It is hardly necessary to point out that the foregoing types of aperiodic substitution are rather unsuitable for practical military usage. Encipherment is slow and subject to error. In some cases encipherment can be accomplished only by single-letter operation. For if the interruptor is a cipher letter the key is interrupted by a letter which cannot be known in advance; if the interruptor is a plain-text letter, while the interruptions can be indicated before encipherment is begun, the irregularities occasioned by the interruptions in keying cause confusion and quite materially retard the enciphering process. In deciphering, the rate of speed would be just as slow in either method. It is obvious that one of the principal disadvantages in all these methods is that if an error in transmission is made, if some letters are omitted, or if anything happens to the interruptor letter, the message becomes difficult or impossible to decryptograph by the ordinary code clerk. Finally, the degree of cryptographic security attainable by most of these methods is not sufficient for military purposes.

SECTION VI

REVIEW OF AUTO-KEY SYSTEMS

Paragraph
The two basic methods of auto-key encipherment 23

23. The two basic methods of auto-key encipherment.—*a.* In auto-key encipherment there are two possible sources for successive key letters: the plain text or the cipher text of the message itself. In either case, the *initial* key letter or key letters are supplied by preagreement between the correspondents; after that the text letters that are to serve as the key are displaced 1, 2, 3, . . . intervals to the right, depending upon the length of the prearranged key.

b. (1) An example of plain-text keying will first be shown, to refresh the student's recollection. Let the previously agreed upon key consist of a single letter, say X, and let the cipher alphabets be direct standard alphabets.

Key	X N O T I F Y Q U A R T E R M A S T E R . .
Plain	N O T I F Y Q U A R T E R M A S T E R . . .
Cipher	K B H B N D O K U R K X V D M S L X V . . .

(2) Instead of having a single letter serve as the initial key, a word or even a long phrase may be used. Thus (using TYPEWRITER as the initial key):

Key	T Y P E W R I T E R| N O T I F Y Q U A R . .
Plain	N O T I F Y Q U A R T E R M A S T E R . . .
Cipher	G M I M B P Y N E I G S K U F Q J Y R . . .

c. (1) In cipher text auto keying the procedure is quite similar. If a single initial key letter is used:

Key	X| K Y R Z E C S M M D W A R D D V O S . . .
Plain	N O T I F Y Q U A R T E R M A S T E R . . .
Cipher	K Y R Z E C S M M D W A R D D V O S J . . .

(2) If a key word is used:

Key	T Y P E W R I T E R| G M I M B P Y N E I . .
Plain	N O T I F Y Q U A R T E R M A S T E R . . .
Cipher	G M I M B P Y N E I Z Q Z Y B H R R V . . .

(3) Sometimes only the last cipher letter resulting from the use of the prearranged key word is used as the key letter for enciphering the auto-keyed portion of the text. Thus, in the last example, the plain text beginning TERMASTER would be enciphered as follows:

Key	T Y P E W R I T E R| I B F W I I A T X . . .
Plain	N O T I F Y Q U A R T E R M A S T E R . . .
Cipher	G M I M B P Y N E I B F W I I A T X O . . .

d. In the foregoing examples, direct standard alphabets are employed; but mixed alphabets, either interrelated or independent, may be used just as readily. Also, instead of the ordinary type of cipher alphabets, one may employ a mathematical process of addition (see par. 40*f* of Special Text No. 166, *Advanced Military Cryptography*) but the difference between the latter process and the ordinary one using sliding alphabets is more apparent than real.

e. Since the analysis of the case in which the cipher text constitutes the auto key is usually easier than that in which the plain text serves this function, the former will be the first to be discussed.

SECTION VII

SOLUTION OF CIPHER-TEXT AUTO-KEY SYSTEMS

Paragraph
Solution of cipher-text auto-keyed cryptograms when known alphabets are employed........ 24
General principles underlying solution of cipher-text auto-keyed cryptograms by frequency analysis........ 25
Frequency distributions required for solution........ 26
Example of solution by frequency analysis........ 27
Example of solution by analysis of isomorphisms........ 28
Special case of solution of cipher-text auto-keyed cryptograms........ 29

24. Solution of cipher-text auto-keyed cryptograms when known alphabets are employed.—*a.* (1) First of all it is to be noted that if the cryptanalyst knows the cipher alphabets which were employed in encipherment, the solution presents hardly any problem at all. It is only necessary to decipher the message beyond the key letter or key-word portion and the initial part of the plain text enciphered by this key letter or key word can be filled in from the context. An example, using standard cipher alphabets, follows herewith:

CRYPTOGRAM

WSGQV OHVMQ WEQUH AALNB NZZMP ESKD

(2) Writing the cipher text as key letters (displaced one interval to the right) and deciphering by direct standard alphabets yields the following:

```
Key............    W S G Q V O H V M Q W E Q U H A A L N B N Z Z M P E S K
Cipher.........  W S G Q V O H V M Q W E Q U H A A L N B N Z Z M P E S K D
Plain..........    W O K F T T O R E G I M E N T A L C O M M A N D P O S T
```

(3) Trial of the word REPORT as the initial word of the message yields an intelligible word as the initial key: FORCE, so that the message reads:

```
Key..................  F O R C E V O H V M Q . .
Cipher...............  W S G Q V O H V M Q . . .
Plain................  R E P O R T T O R E . . .
```

(4) A semiautomatic method of solving such a message is to use sliding normal alphabets and align the strips so that, as one progresses from left to right, each cipher letter is set opposite the letter A on the preceding strip. Taking the letters VMQWEQUHA in the foregoing example, note in Figure 7 the series of placements of the successive strips. Then note how the successive plain-text letters of the word REGIMENT reappear to the left of the successive cipher letters MQWEQUHA.

A	V	H	X	T	X	N	H	O
B	W	I	Y	U	Y	O	I	P
C	X	J	Z	V	Z	P	J	Q
D	Y	K	A	W	A	Q	K	R
E	Z	L	B	X	B	R	L	S
F	A	M	C	Y	C	S	M	T
G	B	N	D	Z	D	T	N	U
H	C	O	E	A	E	U	O	V
I	D	P	F	B	F	V	P	W
J	E	Q	G	C	G	W	Q	X
K	F	R	H	D	H	X	R	Y
L	G	S	I	E	I	Y	S	Z
M	H	T	J	F	J	Z	T	A
N	I	U	K	G	K	A	U	B
O	J	V	L	H	L	B	V	C
P	K	W	M	I	M	C	W	D
Q	L	X	N	J	N	D	X	E
R	M	Y	O	K	O	E	Y	F
S	N	Z	P	L	P	F	Z	E
T	O	A	Q	M	Q	G	A	H
U	P	B	R	N	R	H	B	I
V	Q	C	S	O	S	I	C	J
W	R	D	T	P	T	J	D	K
X	S	E	U	Q	U	K	E	L
Y	T	F	V	R	V	L	F	M
Z	U	G	W	S	W	M	G	N

FIGURE 7.

b. If, as a result of the analysis of several messages (as described in par. 25), mixed primary components have been reconstructed, the solution of subsequent messages may readily be accomplished by following the procedure outlined in *a* above, since in that case the cipher alphabets have become known alphabets.

25. **General principles underlying solution of cipher-text auto-keyed cryptograms by frequency analysis.**—*a.* First of all, it is to be noted in connection with cipher-text auto-keying that repetitions will not be nearly as plentiful in the cipher text as they are in the plain text, because in this system before a repetition can appear two things must happen simultaneously. First, of course, the plain-text sequence must be repeated, and second, one or more cipher-text letters (depending upon the length of the introductory key) immediately before the second appearance of the plain-text repetition must be identical with one or more cipher-text letters immediately before the first appearance of the group. This can happen only as the result of chance. In the following example the introductory key is a single letter, X, and direct standard components are used in the usual Vigenère manner:

Key................ X C K B T M D H N V H L Y . . . . K D K S J M D H N V H L Y
Plain.............. F I R S T R E G I M E N T . . . . T H I R D R E G I M E N T
Cipher............ C K B T M D H N V H L Y R . . . K D K S J M D H N V H L Y R

The repeated plain-text word, REGIMENT, has only 8 letters but the repeated cipher-text group contains 9, of which only the last 8 letters actually represent the plain-text repetition. In order that the word REGIMENT be enciphered by D H N V H L Y R the second time this word appeared in the text it was necessary that the key letter for its first letter, R, be M *both* times; no other key letter will produce the same cipher sequence for the word REGIMENT in this case. Each different key letter for enciphering the first letter of REGIMENT will produce a different encipherment for the word, so that the chances [1] for a repetition in this case are roughly about 1 in 26. This is the principal cause for the reduction in repetitions in this system. If an introductory key of two letters were used, it would be necessary that the two cipher letters immediately before the second appearance of the repeated word REGIMENT be identical with the two cipher letters immediately before the first appearance of the word. In general, then, an n-letter repetition in the cipher text, in this case, represents an $(n-k)$-letter repetition in the plain text, where n is the length of the cipher-text repetition and k is the length of the introductory key.

b. There is a second phenomenon of interest in connection with the cipher-text auto-key method. Let the letter opposite which the key letter is placed (when using sliding components for encipherment) be termed, for convenience in reference, "the base letter." Normally the base letter is the initial letter of the plain component, but it has been seen in preceding texts that this is only a convention. Now when the introductory key is a single letter, if the base letter occurs as a plain-text letter its cipher equivalent is identical with the immediately preceding cipher letter; that is, there is produced a double letter in the cipher text, no matter what the cipher component is and no matter what the key letter happens to be for that encipherment. For example, using the H Y D R A U L I C . . . X Z sequence for both primary components, with H, the initial letter of the plain component as the base letter, and using the introductory key letter X, the following encipherment is produced:

Key	X J O I I F L Y U T T D K K Y C X G
Plain	M A N H A T T A N H I G H J I N K S
Cipher	J O I I F L Y U T T D K K Y C X G L

Note the doublets II, TT, KK. Each time such a doublet occurs it means that the second letter represents H_p, which is the base letter in this case (initial letter of plain component). Now if the base letter happens to be a high-frequency letter in normal plain text, for example the letter E, or T, then the cipher text will show a large number of doublets; if it happens to be a low-frequency letter the cipher text will show very few doublets. In fact, the number of doublets will be directly proportional to the frequency of the base letter in normal plain text. Thus, if the cryptogram contains 1,000 letters there should be about 72 occurrences of doublets if the base letter is A, since in 1,000 letters of plain text there should be about 72 A's. Conversely, if a cryptogram of 1,000 letters shows about 72 doublets, the base letter is likely to be A; if it shows about 90, it is likely to be T, and so on. Furthermore when a clue to the identity of the base letter has been obtained in this manner, it is possible immediately to insert the corresponding plain-text letter throughout the text of the message. The distribution of this letter may not only serve as a check (if no inconsistencies develop) but also may lead to the assumption of values for other cipher letters.

c. When the introductory key is 2 letters, then this same phenomenon will produce groups of the formula ABA, where A and B may be any letters, but the first and third must be identical. The occurrence of patterns of this type in this case indicates the encipherment of the base letter.

---

[1] If all the cipher letters appeared with equal frequency the chances would be exactly 1 in 26. But certain letters appear with greater frequency because some plain-text letters are much more frequent than others.

d. The phenomena noted above can be used to considerable advantage in the solution of cryptograms of this type. For instance, if it is known that the ordinary Vigenère method of encipherment is used ($\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{c/2}$), then the initial letter of the plain component is the base letter. If, further, it is known that the plain component is the normal direct sequence, then the base letter is A and a word such as BATTALION will be enciphered by a group having the formula $\underline{\text{AA}}\text{B}\underline{\text{CC}}\text{DEFG}$. If the plain component is a mixed sequence and happens to start with the letter E, then a word such as ENEMY would be enciphered by a sequence having the formula $\underline{\text{AA}}\underline{\text{BB}}\text{CD}$.[2] Sequences such as these are, of course, idiomorphic and if words yielding such idiomorphisms are frequent in the text there will be produced in the latter several or many cases of isomorphism. When these are analyzed by the principles of indirect symmetry of position, a quick solution may follow.

e. A final principle underlying the solution of cipher-text auto-keyed cryptograms remains to be discussed. It concerns the nature of the frequency distributions required for the analysis of such cryptograms. This principle will be set forth in the next paragraph.

26. Frequency distributions required for solution.—*a.* Consider the message given in paragraph 23*c* (1). It happens that the letter R_c occurs twice in this short message and, because of the nature of the cipher-text auto-keying method, this letter must also appear twice in the key. Now it is obvious that all plain-text letters enciphered by key letter R_k will be in the same cipher alphabet; in other words, if the key text is "offset" one letter to the right of the cipher text, *then every cipher letter which immediately follows an* R_c *in the cryptogram will belong to the same cipher alphabet*, and this alphabet may be designated conveniently as the R cipher alphabet. Now if there were sufficient text, so that there were, say, 30 to 40 R_c's in it, then a frequency distribution of the letters immediately following the R_c's will exhibit monoalphabeticity. What has been said of the letters following the R_c's applies equally well to the letters following all the other letters of the cipher text, the A_c's, B_c's, C_c's, and so on. In short, if 26 distributions are made, one for each letter of the alphabet, showing the cipher letter immediately succeeding each different letter of the cipher text, then the text of the cryptogram can be allocated into 26 uniliteral, monoalphabetic frequency distributions which can be solved by frequency analysis, providing there are sufficient data for this purpose.

b. The foregoing principle has been described as pertaining to the case when the introductory key is a single letter, that is, when the key text is "offset" or displaced but one interval to the right of the cipher text. But it applies equally to cases wherein the key text is offset more than one interval, provided the frequency distributions are based upon the proper interval, as determined by the displacement due to the length of the introductory key. For instance, suppose the introductory key consists of two letters, as in the following example:

Key text.................. $\underline{\text{X Z}}$| M R H F H G F N Q R X O M R M V W E E
Plain text................ R E L I A B L E I N F O R M A T I O N . .
Cipher text............... M R H F H G F N Q R X O M R M V W E E . .

The key text in this case is offset two intervals to the right of the cipher text and, therefore, frequency distributions made by taking the cipher letters one interval to the right of a given cipher letter, each time that letter occurs, will not be monoalphabetic because some letter not related at all to the given cipher letter is the key letter for enciphering the letter one interval to the right of the latter. For example, note the three R_c's in the foregoing illustration. The first R_c is followed by H_c, representing the encipherment of L_p by M_k; the second R_c is followed by X_c, representing the encipherment of F_p by Q_k; the third R_c is followed by M_c, representing the encipherment of A_p by M_k. The three cipher letters H, X, and M are here entirely unrelated and do not belong to the same cipher alphabet because they represent encipherments by three different key letters. On the other hand, the cipher letters two intervals to the right of the R_e's, viz, F, O, and V, are in the same cipher alphabet because these cipher letters are the results of enciphering plain-text letters I, O, and T, respectively, by the *same* key letter, R. It is obvious, then, that when the introductory key consists of two letters and the key text is displaced two intervals to the right of the cipher text, the proper frequency distributions for monoalphabeticity will be based upon the letter at the second interval to the right of each cipher letter. Likewise, if the introductory key consists of three letters and the key text is displaced three intervals to the right of the cipher text, the distributions must be based upon the third interval, and so on, in each case the interval used corresponding to the amount of displacement between key text and cipher text.

---

[2] Six letters are shown because the idiomorphism in this case extends over that many letters.

c. Conversely, in solving a problem of this type, when the length of the introductory key and therefore the amount of displacement are not known, the appearance of the frequency distributions based upon various intervals after each different cipher letter will disclose this unknown factor, since only one set of distributions will exhibit monoalphabeticity and the interval corresponding to that set will be the correct interval.

d. Application of these principles will now be made, using a specific example.

27. Example of solution by frequency analysis.—*a.* It will be assumed that previous studies have disclosed that the enemy is using the cipher-text auto-key system described. It will be further assumed that these studies have also disclosed that (1) the introductory key is usually a single letter, (2) the usual Vigenère method of employing sliding primary components is used, (3) the plain component is usually the normal direct sequence, the cipher component a mixed sequence which changes daily. The following cryptograms, all of the same date, have been intercepted:

Message I

```
I J X W X   E E C D A   C N Q E T   U K N M V   D I W P P
Q Z S X D   H I F E L   N N J J I   D I V E Y   G T C Z M
E H H L M   R V C U R   G D I E Q   S G T A R   J J Q Q Y
C A R P H   M G L D Y   F Y T C D   G Y F K R   F K S E T
T D I Q K   K M L T U   R Q G G N   K M K I X   J X W K A
O K N T B   T Z J O Q   Y S C D I   D G E T X   G X X X X
```

Message II

```
G R V R M   Z W K X G   W P C K K   R M X A N   J C C X U
R T N J U   A K O B L   N L M W K   Y Y Z J U   C S U H F
F H I J A   Q B M L T   P U R R S   U E Q E V   Z E Y G C
F F N F I   B W N Y S   T C E T P   D G T T Z   R R Q H Q
A O O X D   B U Y N K   L B W C D   G G K X X
```

Message III

```
R W K A O   L T C J M   Z D K V U   J C D D Y   B Z E L M
M W T Q O   H Q V G X   C H O L M   W V G R K   I B R X D
L A Q Y U   K I R O Z   T Q Y U X
```

Message IV

XJJPM LTZKX ECAQZ NTTOC ONDUC
TUTCV GRJPF FDIPP DIXCE SETWW
SUMUJ CSLGX HXMOZ EKAQI SUAOX

Message V

GISUH WZHST TZOID DHOOV NBTJG
XCTBS FKIRH MMVYM IIVUU CZMJE
HAGIE WMEHH LMWKY PPDQZ GBOIW
PSFAJ UQZHZ MTFHZ MLACZ ROVDI
WPVIB OBCCX NNDGI ESJOC KBJHQ
MUZEL YOOVU JWKIE IBBCZ AJIEF
FORSA JLNQM BQXXX

Message VI

TBJPA ARYYP VHIDI TUXNJ MXGSS
BDAQY MMTTF UUNMG QPUXM OVUYE
CECZM MWOHC FOBHV NKAZC KMXXX

Message VII

TBJPA QAAZT RXALX FKKME IAABD
SFTQT CJJGJ OVMRG LVWTT JUAWL
XUKTX GGBOX MXDID SPBSF LYZKC
FXXXX

b. A distribution table of the type described in paragraph 25*e* is compiled and is shown as Figure 8 below. In making these distributions it is simple to insert a tally in the appropriate cell in the pertinent horizontal line of the table, to indicate the cipher letter which immediately follows each occurrence of the letter to which that line applies. Obviously, the best method of compiling the data is to handle the text digraphically, taking the first and second letters, the second and third, the third and fourth, and so on, and distributing the final letters of the digraphs in a quadricular table. The distribution merely takes the form of tally marks, the fifth being a diagonal stroke so as to totalize the occurrences visibly.

SECOND LETTER

FIRST LETTER	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	///	/	//				/			///	/	/		/	////		𝍸/	///					/			//
B		/	/	///				/		///		/	/		////		/	/	//	///	/		//			/
C	//		//	𝍸	///	///		/		//	///			/					//	/	/	/		///		////
D	//	/		//		/	𝍸	//	𝍸𝍸		/	/					/		//						//	
E			////		/	/		///	//		/	///					//		//	𝍸		/	/		//	
F	/			/	/	////		//	/	/	////	/		/	//					/	/				/	
G		//	/	/	/		///		///	/	/	//		/			/	///	/	///			/	//	/	
H	/		/			/		//	///			//	/		//		///		/			/	/	/		//
I	/	////		𝍸	𝍸	/			/	//						/	/	//	//	/		//	///	//		
J	/		////		/		/	/	//	////		/	///		///	////	/				///		/	//		
K	////	/	/						𝍸		///	/	////	//	/			//	/	/		/		//	//	
L	//	/		/			/						𝍸	///						////		/		//	//	
M		///			///		///		/	/	/	////	////		//			//		//	//	//	𝍸/	///		//
N		/		/		/				////	///	/	//	//			//			/					/	
O		///	/					//	//		/	//			///		/	/				𝍸		//		///
P	//	/	/	///		/		/					/			///	/		/		//	//				
Q	//	/			//		/	/	/		/		//		/	/	/		/	/		/			𝍸	////
R						/	//	/		//	/		//		//	/	//	//	//	/		//	/	//	/	
S	/	/	/		//	////	/			/		/				/			/	//	𝍸			/		
T	/	////	𝍸/	/		//				//				/		//	///	/		𝍸	////		/	//		////
U	///		///		/			//		///	///		/	/			/	////		/	//			//	//	/
V			/	//	/		///	/	/				/	//				/			////		/		/	/
W			/								𝍸/	/	/	/	/	////			/	//		/	/	//		/
X	//		///	////	//	/	////	/		//			////	//							//		//			
Y		/	/		/	//	//						//	/	/	//			//	/	//				//	//
Z	/		/	/	////		/	//		//	//		𝍸/		/			//	/	//			/			

FIGURE 8.

c. The individual frequency distributions give every appearance of being monoalphabetic, which checks the assumption that the enemy is still employing the same system. The total number of letters of text (excluding the final X's) is 680. If the base letter is A then there should be approximately $680 \times 7.2\% = 49$ cases of double letters in the text. There are actually 52 such cases, which checks quite well with expectancy. The letter A is substituted throughout the text for the second letter of each doublet.

d. The following sequence is noted:

Message V, line 1.......... G I S U H W Z H S $\overline{\text{T T}}$ Z O I $\overline{\text{D D}}$ H $\overline{\text{O O}}$ V N B T J G
. . . A . . . . A . . A . . .

Assume that the sequence DDHOOVNBT represents BATTALION. Then the frequency of H_c in the D cipher alphabet should be high, since $H_c = T_p$. The H has only 2 occurrences. Likewise, the frequency of O_c in the H alphabet ($=T_p$) should be high; it is also only 2. The frequency of V in the O alphabet should be medium or low, since it would equal L_p; it is 5, which is too high. The rest of the letters of the assumed word are similarly checked against the appropriate frequency distributions, with the result that, on the whole, the assumption that the DDHOOVNBT sequence represents BATTALION does not appear to be warranted. Similar attempts are made at other points in the text, with the same or other probable words. Some of these attempts may have to be carried to the point where the placement of values in the tentative cipher component leads to serious inconsistencies. Finally, attention is fixed upon the following sequence:

Message VI, line 2........ B D A Q Y $\overline{\text{M M}}$ $\overline{\text{T T}}$ F $\overline{\text{U U}}$ N M G . . .
. . . A . A . . A . . .

The word $\overset{\text{MMTTFUUNMG}}{\text{AVAILABLE}}$ is assumed. The appropriate frequency distributions are consulted to see how well the actual individual frequencies correspond to the expected ones.

Alphabet	Assumed		Frequency		Approximation
	θ_c	θ_p	Expected	Actual	
M	T	V	Low	2	Fair
T	F	I	High	2	Fair
F	U	L	Medium	1	Good
U	N	B	Low	1	Good
N	M	L	Medium	2	Fair
M	G	E	High	3	Fair

The assumption cannot be discarded just yet. Let the values derivable from the assumption be inserted in their proper places in a cipher component, and, using the latter in conjunction with a normal direct sequence as the plain component, let an attempt be made to find corroboration for these values. The following placements may be made:

Plain..............	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.............	M			F	G										U	N						T				

The letter M_c appears twice in the cipher sequence and when this partially reconstructed cipher component is tested it is found that the value $L_p(N_k)=M_c$ is corroborated. Having the letters M, F, G, U, N, and T tentatively placed in the cipher component, it is possible to insert certain plain-text values in the text. For example, in the M alphabet, $F_c=D_p$, $G_c=E_p$, $U_c=O_p$, $N_c=P_p$, $T_c=V_p$. In the F alphabet, $G_c=B_p$, $U_c=L_p$, $N_c=M_p$, $T_c=S_p$, $M_c=X_p$. The other letters yield additional values in the appropriate alphabets. The plain-text values thus obtainable are inserted in the cipher text. No inconsistencies appear and, moreover, certain "good" digraphs are brought to light. For instance, note what happens here:

	Key..........	.	U	Q	Z	H	Z	M	T	F	H	Z	M	L	A	C	Z
Message V, line 4..............	Cipher........	U	Q	Z	H	Z	M	T	F	H	Z	M	L	A	C	Z	.
	Plain..........							V	I								

Now if the letter H can be placed in the cipher component, several values might be added to this partial decipherment. Noting that F and G are sequent in the cipher component, suppose H follows G therein. Then the following is obtained:

	Key..........	.	U	Q	Z	H	Z	M	T	F	H	Z	M	L	A	C	Z
Message V, line 4..............	Cipher........	U	Q	Z	H	Z	M	T	F	H	Z	M	L	A	C	Z	.
	Plain..........							V	I	C							

Suppose the VIC is the beginning of VICINITY. This assumption permits the placement of A, C, L, and Z in the cipher component, as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	M	A		F	G	H			L					Z	U	N						T				C

These additional values check in very nicely and presently the entire cipher component is reconstructed. It is found to be as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	M	A	B	F	G	H	J	K	L	Q	S	V	X	Z	U	N	D	E	R	W	O	T	Y	P	I	C

The key phrase is obviously UNDERWOOD TYPEWRITER COMPANY. All the messages now may be deciphered with ease. The following gives the letter-for-letter decipherment of the first three groups of each message:

I (Introductory key: K)

Key	K I J X W	X E E C D	A C N C Q	. . .
Cipher	T J X W X	E E C D A	C N Q E T	. . .
Plain	R I G H T	F A I R L	Y Q U I E	. . .

II (Introductory key: E)

Key	E G R V R	M Z W K X	G W P C K	. . .
Cipher	G R V R M	Z W K X G	W P C K K	. . .
Plain	N O T H I	N G O F S	P E C I A	. . .

III (Introductory key: R)

Key	R R W K A	O L T C J	M Z D K V	. . .
Cipher	R W K A O	L T C J M	Z D K V U	. . .
Plain	A B O U T	O N E H U	N D R E D	. . .

IV (Introductory key: J)

Key	J X J J P	M L T Z K	X E C A Q	. . .
Cipher	X J J P M	L T Z K X	E C A Q Z	. . .
Plain	G U A R D	I N S U F	F I C I E	. . .

V (Introductory key: E)

Key	E G I S U	H W Z H S	T T Z O I	. . .
Cipher	G I S U H	W Z H S T	T Z O I D	. . .
Plain	N U M E R	O U S F L	A S H E S	. . .

VI (Introductory key: B)

Key	B T B J P	A A R Y Y	P V H I D	. . .
Cipher	T B J P A	A R Y Y P	V H I D I	. . .
Plain	T H E R E	A R E A B	O U T S I	. . .

VII (Introductory key: B)

Key	B T B J P	A Q A A Z	T R X A L	. . .
Cipher	T B J P A	Q A A Z T	R X A L X	. . .
Plain	T H E R E	I S A M I	X U P H E	. . .

e. In the foregoing example the plain component was the normal direct sequence, so that with the Vigenère method of encipherment the base letter is A. If the plain component is a mixed sequence, the base letter may no longer be A, but in accordance with the principle set forth in paragraph 25*b*, the frequency of doublets in the cipher text will correspond with the frequency of the base letter as a letter of normal plain text. If a good clue as to the identity of this letter is afforded by the frequency of doublets in the cipher text, the insertion of the corresponding base letter in the plain text will lead to further clues. The solution from there on can be handled along the lines indicated above.

28. **Example of solution by analysis of isomorphisms.**—*a.* It was stated in paragraph 25*d* that in cipher-text auto-keying the production of isomorphs is a frequent phenomenon and that analysis of these isomorphs may yield a quick solution. An example of this sort will now be studied.

b. Suppose the following cryptograms have been intercepted:

1

```
U S Y P W   T R X D I   M L E X R   K V D B D   D Q G S U   N S F B O
B E K V B   M A M M O   T X X B W   E N A X M   Q L Z I X   D I X G Z
P M Y U C   N E V V J   L K Z E K   U R C N I   F Q F N N   Y G S I J
T C V N I   X D D Q Q   E K K L R   V R F R F   X R O C S   S J T B V
E F A A G   Z R L F D   N D S C D   M P B B V   D E W R R   N Q I C H
A T N N B   O U P I T   J L X T C   V A O V E   Y J J L K   D M L E G
N X Q W H   U V E V Y   P L Q G W   U P V K U   B M M L B   O A E O T
T N K K U   X L O D L   W T H C Z   R
```

2

```
B I I B F   G R X L G   H O U Z O   L L Z N A   M H C T Y   S C A A T
X R S C T   K V B W K   O T G U Q   Q F J O C   Y Y B V K   I X D M T
K T T C F   K V K R O   B O E P L   Q I G N R   I Q O V J   Y K I P H
J O E Y M   R P E E W   H O T J O   C R I I X   O Z E T Z   N K
```

3

```
H A L O Z   J R R V M   M H C V B   Y U H A O   E O V A C   Q V V J L
K Z E K U   R F R F X   Y B H A L   Z O F H M   R S J Y L   A P G R S
X A G X D   M C U N X   X L X G Z   J P W U I   F D B B Y   P V F Z N
B J N N B   I T M L J   O O S E A   A T K P B   Y
```

c. Frequency distributions are made, based upon the 2d letters of pairs, as in the preceding example. The result is shown in the table in figure 9. The data in each distribution are relatively scanty and it would appear that the solution is going to be a rather difficult matter.

SECOND LETTER

FIRST LETTER

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	//		/		/		//					//	//		//	/				///				/		
B		/		/	/	/		/	//	/			//		////							///	//		///	
C	/			/		/		/						//			/	/	/	//	/	///			/	/
D		//		//	/				//			/	////	/			//		/							
E	/				/	/	/				////			/	//	/				/		//	//	/	//	
F	/	/		//			/	/		/	/			/			/	//						//		/
G								/						//				//	//		/		/	/		///
H	////		///							/			/		//						/					
I			/			//	/	/	/				/			/	/			//				𝍸		
J										/		////		/	////	/		/		//					//	
K				/					//		//	/			/	/		/		/	////	////				//
L	/	/			//	/	/			/	///	/			///		//	/					/	//		///
M	/		/					//				////	///		/	/	/	//		/					/	
N	//	///		/	/				//		//			///			/	/	/					//	/	
O	/	//	///	/	///	/						/			/				/	////	//	///				//
P		//			/		/	/	/			//	/									//	//			
Q					/	//	//		//			/			/		//					/	/			
R			/			////			//		/	/		/	//	/		//	///			//		//		
S			///		/	/			/	//									/		/			/	/	
T		/	///				/	/		//	///		/	//				/		//				//	/	/
U		/	/					/	/					//		//	/	//	/			/		/		/
V	//	///		//	///	/				///	///		/	/				/				//			/	
W		/			/			//			/							/		//	//					
X	/	/		𝍸			//					///	/		/		/	///		/				//	/	
Y		//					/			/	/	/	/			///			/		//				/	
Z					///				/	//				///	//	/		//								

FIGURE 9.

d. However, before becoming discouraged too quickly, a search is made throughout the text to see if any isomorphs are present. Fortunately there appear to be several of them. Note the following:

Message 1 .... (1) . D B D D Q G S U N S F B O B E K . . .
(2) . N E V V J L K Z E K U R C N I F . . .
(3) . T N K K U X L O D L W T H C Z R | end of message
Message 2 .... (4) . C R I I X O Z E T Z N K | end of message
Message 3 .... (5) . C Q V V J L K Z E K U R F R F X . . .

First, it is necessary to delimit the length of the isomorphs. Isomorph (2) shows that the isomorphism begins with the doubled letters. For there is an E before the V V in that case and also an E within the isomorph; if the phenomenon included the E, then the letter immediately before the D D in the case of isomorph (1) would have to be an N, to match its homolog, E, in isomorph (2), which it is not. Corroborating data are given by isomorphs (3), (4), and (5) in this respect. Hence, we may take it as established that the isomorphism begins with the doubled letters.

As for the end of the isomorphism, the fact that isomorphs (2) and (5) are the same for 10 letters seems to indicate that that is the length of the isomorphism. The fact that message 2 ends 2 letters after the last "tie-in" letter, Z, corroborates this assumption. It is at least certain that the isomorphism does not extend beyond 11 letters because the recurrence of R in isomorph (5) is not matched by the recurrence of R in isomorph (2), nor by the recurrence of T in isomorph (3). Hence it may be assumed that the isomorphic sequence is probably 10 letters in length, possibly 11. But to be on safe ground it is best to proceed on the 10-letter basis.

e. Applying the principles of indirect symmetry to the superimposed isomorphs, partial chains of equivalents may be constructed and it happens in this case that practically the entire primary component may be established. Let the student confirm the fact that the following sequence may be derived from the data given:

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
T E Z K R . I V F .  .  .  Q  .  W  G  .  N  U  S  B  X  J  D  O  L
```

The only missing letters are A, C, H, M, P, and Y. By use of the nearly complete sequence on the text it will be possible to place these 6 letters in their positions in the cipher component. Or, if a keyword-mixed sequence is suspected, then the sequence which was reconstructed may be merely a decimation of the original primary sequence. By testing the partial sequence for various intervals, when the seventh is selected the following result is obtained:

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
T V W X Z . . D R .  U  L  I  .  B  E  F  G  J  K  .  N  O  .  Q  S
```

The sequence is obviously based on the keyword HYDRAULIC, and the complete primary cipher component is now available. The plain component is then to be reconstructed. A word must be assumed in the text.

f. A good probable word to assume for the 10-letter repetition found in messages 1 and 3 is ARTILLERY. This single assumption is sufficient to place 7 letters in the plain component. Thus:

```
Key.............. . . . V V J L K Z E K U R . . .
Plain............ . . . A R T I L L E R Y . . . .
Cipher........... . . V V J L K Z E K U R . . . .
```

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
A . . . E . . . I .  .  L  .  .  .  .  .  R  .  T  .  .  .  .  Y  .
```

These few letters are sufficient to indicate that the plain component is probably the normal direct sequence. A few minutes testing proves this to be true. The two components are therefore:

```
Plain............ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher........... H Y D R A U L I C B E F G J K M N O P Q S T V W X Z
```

With these two components at hand, the decipherment of the messages now becomes a relatively simple matter. Assuming a single-letter introductory key, and trying the first five groups of message 1 the results are as follows:

```
Key.............. ? U S Y P  W T R X D  I M L E X  R K V D B  D D Q G S  . . .
Cipher........... U S Y P W  T R X D I  M L E X R  K V D B D  D Q G S U  . . .
Plain............ ? P H R F  Y I V E F  I R E O F  L I G H T  A R T I L  . . .
```

It is obvious that an introductory key of more than one letter was used, since the first few letters yield unintelligible text; but it also appears that the last cipher letter of the introductory key was used as the introductory key letter for enciphering the subsequent auto-keyed portion of the text (see par. 23*c*(3)). However, assuming that the IVE before the word FIRE is the ending of the first word of the plain text, and that this word is INTENSIVE, the introductory key word is found to be WICKER. Thus:

```
Key.......... W I C K E R|T R X D I M L E X R K V D B D D Q G S . . .
Plain........ I N T E N S I V E F I R E O F L I G H T A R T I L . . .
Cipher....... U S Y P W T R X D I M L E X R K V D B D D Q G S U . . .
```

The beginnings of the other two messages are recoverable in the same way and are found to be as follows:

```
Key............... P R O M I S E|R X L G H O U Z O . .
Plain............. R E Q U E S T V I G O R O U S . . .
Cipher............ B I I B F G R X L G H O U Z O . . .

Key............... C H A R G E D|R R V M M H C V B . . .
Plain............. S E C O N D B A T T A L I O N . . . .
Cipher............ H A L O Z J R R V M M H C V B . . . .
```

g. The example solved in the foregoing subparagraphs offers an important lesson to the student, insofar as it teaches him that *he should not immediately feel discouraged when confronted with a problem presenting only a small quantity of text and therefore affording what seems at first glance to be an insufficient quantity of data for solution.* For in this example, while it is true that there are insufficient data for analysis by simple principles of frequency, it turned out that solution was achieved *without any recourse to the principles of frequency of occurrence.* Here, then, is one of those interesting cases of substitution ciphers of rather complex construction which are solvable without any study whatsoever of frequency distributions. Indeed, it will be found to be true that in more than a few instances the solution of quite complicated cipher systems may be accomplished not by the application of the principles of frequency, but by recourse to inductive and deductive reasoning based upon other considerations, even though the latter may often appear to be very tenuous and to rest upon quite flimsy supports.

29. Special case of solution of cipher-text auto-keyed cryptograms.—*a*. Two messages with identical plain texts enciphered according to the method of paragraph 23 *c* (3) by initial key words of different lengths and compositions can be solved very rapidly by reconstructing the primary components. The *cryptographic texts of such messages will be isomorphic after the initial key-word portions.* Note the two following superimposed messages, in which isomorphism between the two cryptograms is obvious after their 6th letters:

```
1. T S B J S   K B N L O   C F H A Z   L W J A M   B N F N S   M V J R E
2. B K K M J   X Y C X B   H R P V O   X M U V I   Y C R C G   I K U T D

1. H F P R X   C P C R R   E H F M U   H R A X C   N F D U B   A T F Q R
2. P R E T N   H E H T T   D P R I W   P T V N H   C R S W Y   V J R F T
```

Starting with any pair of superimposed letters (beginning with the 7th pair), chains of equivalents are constructed:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1................	Z	O	B	Y	.	.	.							
2................	L	X	N	C	H	P	E	D	S	G	.	.	.	.
3................	Q	F	R	T	J	U	W	M	I	.	.	.		
4................	A	V	K	.	.	.								

By interpolation, these partial sequences may be united into the key-word sequence:

H Y D R A U L I C B E F G J K M N O P Q S T V W X Z

b. The initial key words and the plain texts may now be ascertained quite easily by deciphering the messages, using this primary component slid against itself. It will be found that the initial key word for the 1st message is PENCE, that for the 2d is LATERAL. The reason that the cryptographic texts are isomorphic beyond the initial key word portions is, of course, that since the text beyond the key word is enciphered auto-key fashion by the preceding cipher letter the letters before the last letter of the key have no effect upon the encipherment at all. Hence two messages of identical text cannot be other than isomorphic after the initial key-word portions.

c. The foregoing solution affords a clue to the solution of cases in which the texts of two or more messages are not completely identical but are in part identical because they happen to have similar beginnings or endings, or contain nearly similar information or instructions. The progress in such cases is not so rapid as in the case of messages with wholly identical texts because much care must be exercised in blocking out the isomorphic sequences upon which the reconstruction of the primary components will be based.

d. (1) In the foregoing cases, the primary components used to encipher the illustrative messages were identical mixed sequences. If nonidentical components are employed, the cryptograms present an interesting case for the application of a principle pointed out in a preceding text.[4]

(2) Suppose that the three messages of paragraph 27*b* had been enciphered by using a plain component different from the mixed component. The encipherments of the word ARTILLERY would still yield isomorphic sequences, from which, as has been noted, the reconstruction of the cipher component can be accomplished.

(3) Having reconstructed the cipher component (or an equivalent) the latter may be applied to the cipher text and a "decipherment" obtained. In this process *any* sequence of 26 letters may be used as the plain component and even the normal sequence A . . . Z may be employed for this purpose. The word decipherment in the next to the last sentence is enclosed by quotation marks because the letters thus obtained would not yield plain text, since the real or an equivalent plain component has not yet been found. Such "deciphered" text may be termed *spurious* plain text. *But the important thing to note is that this text is now monoalphabetic and may be solved by the simple procedure usually employed in solving a monoalphabetic cipher produced by a single mixed alphabet.* Thus, a polyalphabetic cipher may be converted to monoalphabetic terms and the problem much simplified. In other words, here is another example of the situations in which the principle of conversion into monoalphabetic terms may be applied with gratifying success. It is also an example of the dictum that the use of two differently mixed primary components does not really give much more security than does a mixed component sliding against itself or against the normal sequence.

---

[4] *Military Cryptanalysis, Part II*, par. 45*g*.

e. (1) If the auto-key method shown in paragraph 23*c* (2) had been employed in enciphering the two identical texts above, the solution would, of course, have been a bit more difficult. To illustrate such a case, let the two texts be enciphered by key words of the same lengths but different compositions: PENCE and LATER. Thus:

No. 1

Key	P E N C E	T S B J S	M M N R U	L P U I H	J B T X F	I N N R M
Plain	R E Q U E	S T I N F	O R M A T	I O N O F	S I T U A	T I O N I
Cipher	T S B J S	M M N R U	L P U I H	J B T X F	I N N R M	D W I Q V
Key	D W I Q V	P C K A O	D P A Z O	B C M R I	A F N W O	G L I H T
Plain	N F I F T	E E N T H	I N F A N	T R Y S E	C T O R A	T O N C E
Cipher	P C K A O	D P A Z O	B C M R I	A F N W O	G L I H T	I W W C U

No. 2

Key	L A T E R	B K K M J	R B T U X	S G E B Q	Y R H H A	T E T U C
Plain	R E Q U E	S T I N F	O R M A T	I O N O F	S I T U A	T I O N I
Cipher	B K K M J	R B T U X	S G E B Q	Y R H H A	T E T U C	N O G T M
Key	N O G T M	L D Q L E	N G B Y E	W D S U H	P U T Z E	H H G D K
Plain	N F I F T	E E N T H	I N F A N	T R Y S E	C T O R A	T O N C E
Cipher	L D Q L E	N G B Y E	W D S U H	P U T Z E	H H G D K	T O D E X

(2) Now let the two cryptograms be superimposed and isomorphisms be sought. They are shown underlined below:

1	T S B J S	M M N R U	L P U I H	J B T X F	I N N R M	D W I Q V
2	B K K M J	R B T U X	S G E B Q	Y R H H A	T E T U C	N O G T M
1	P C K A O	D P A Z O	B C M R I	A F N W O	G L I H T	I W W C U
2	L D Q L E	N G B Y E	W D S U H	P U T Z E	H H G D K	T O D E X

It will be noted that the intervals between isomorphic superimposed pairs show a constant factor of 5, indicating a 5-letter intial key word.

(3) A reconstruction diagram for the pairs beyond the first five letters is established, based upon this interval of 5, and is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	P	W		N			H		T	Y	D	S	R			L	I	O					F	G		
2	X	R	D			U						H	B	E		G		W					O	P		
3	B	K		I		N	O		G		Q		S	T		W	X	C		H	E		D	R		
4	L	F	E	A				D	B		N	C		P		S	T	U		W			Z	H		Y
5	W	D		T		A	U	Q	H		I		C	B	E	F	G			K	X	M	N	O		

The equivalent sequence A W N B D T K I H Q G U X O E R V M C Y S J L Z P F is established by indirect symmetry; from this, by decimation on the eleventh interval, the HYDRAULIC . . . XZ component is recovered.

(4) It will be noted that the foregoing case, in which the initial key words for the two cryptograms are of the same length, is only a special application of the method set forth in paragraph 44 of Military Cryptanalysis, Part II. But if the key words were of different lengths, the method set forth in paragraph 45 of the text referred to would be applicable. No example is deemed necessary, since no new principles are involved.

SECTION VIII

SOLUTION OF PLAIN-TEXT AUTO-KEY SYSTEMS

Paragraph
Preliminary remarks on plain-text auto-keying 30
Solution of plain-text auto-keyed cryptograms when the introductory key is a single letter 31
Example of solution by the probable-word method 32
Concluding remarks on the solution of auto-key systems 33

30. Preliminary remarks on plain-text auto-keying.—*a*. If the cipher alphabets are unknown sequences, plain-text auto-keying gives rise to cryptograms of more intricate character than does cipher-text auto-keying, as has already been stated. As a cryptographic principle it is very commonly encountered as a new and remarkable "invention" of tyros in the cryptographic art. It apparently gives rise to the type of reasoning to which attention has been directed once before and which was then shown to be a popular delusion of the uninitiated. The novice to whom the auto-key principle comes as a brilliant flash of the imagination sees only the apparent impossibility of penetrating a secret which enfolds another secret. His reasoning runs about as follows: "In order to read the cryptogram, the would-be solver must, of course, first know the key; but the key does not become known to the would-be solver until he has read the cryptogram and has thus found the plain text. Since this is reasoning around a circle, the system is indecipherable." How unwarranted such reasoning really is in this case, and how readily the problem is solved, will be demonstrated in the next few paragraphs.

b. A consideration of the mechanics of the plain-text auto-key method discloses that a repetition of n letters in the plain text will produce a repetition of $(n-k)$ letters in the cipher text, where n represents the length of the repetition and k the length of the introductory key. Therefore, when the introductory key consists of a single letter there will be as many repetitions in the cipher text as there are in the plain text, except for true digraphic repetitions, which of course disappear. But on the other hand some "accidental" digraphic repetitions are to be fairly expected, since it can happen that two different plain-text pairs, enciphered by different key letters, will produce identical cipher equivalents. Such accidental repetitions will happen less frequently, of course, in the case of longer polygraphs, so that when repetitions of 4 or more letters are found in the cipher text they may be taken to be true or causal repetitions. It is obvious that in studying repetitions in a cryptogram of this type, when the introductory key is a single letter, a 5-letter repetition in the cipher text, for example, represents a 6-letter word, or sequence repeated in the plain text. When the introductory key is k letters in length then an n-letter repetition represents an $(n+k)$-letter repetition in the plain text.

c. The discussion will, as usual, be divided into two principal cases: (1) when the cipher alphabets are known and (2) when they are unknown. Under each case there may be an introductory key consisting of a single letter, a word, or a short phrase. The single-letter initial key will be treated first.

31. Solution of plain-text auto-keyed cryptograms when the introductory key is a single letter.—*a*. Note the following plain-text auto-keyed encipherment of such commonly encountered plain-text words as COMMANDING, BATTALION, and DIVISION, using two identical primary components, in this case direct standard alphabets:

(1)	Key text	. B A T T A L I O N	Key text	. D I V I S I O N	(2)
	Plain text	B A T T A L I O N .	Plain text	D I V I S I O N .	
	Cipher	. B T M T L T W B .	Cipher	. L D D A A W B .	
(3)	Key text	. C O M M A N D I N G	Key text	. C A P T A I N	(4)
	Plain text	C O M M A N D I N G .	Plain text	C A P T A I N .	
	Cipher	. Q A Y M N Q L V T .	Cipher	. C P I T I V .	

These characteristics may be noted:[1]

(1) The cipher equivalent of A_p is the plain-text letter which immediately precedes A_p. (See the two A's in BATTALION, in example 1 above.)

(2) A plain-text sequence of the general formula ABA yields a doublet as the cipher equivalent of the final two letters. (See IVI or ISI in DIVISION, example 2 above.)

(3) Every plain-text trigraph having A_p as its central letter yields a cipher equivalent the last two letters of which are identical with the initial and final letters of the plain-text trigraph. (See MAN in COMMANDING, example 3 above.)

(4) Every plain-text tetragraph having A_p as the initial and the final letter yields a cipher equivalent the second and fourth letters of which are identical with the second and third letters of the plain-text tetragraph, respectively. (See APTA in CAPTAIN, example 4 above; also ATTA in BATTALION, example 1.)

b. (1) From the foregoing characteristics and the fact that a repetition of a sequence of *n* plain-text letters will yield, in the case of a 1-letter introductory key, a repetition of a sequence of *n-l* cipher letters, it is obvious that the simplest method of solving this type of cipher is that of the probable word. Indeed, if the system were used for regular traffic it would not be long before the solution would consist merely in referring to lists of cipher equivalents of commonly used words (as found from previous messages) and searching through the messages for these cipher equivalents.

(2) Note how easily the following message can be solved:

B E C J I   B T M T L   T W B P Q   A Y M N Q   H V N E T   W A A L C . . .

Seeing the sequence BTMTLTWB, which is on the list of equivalents in *a* above (see example 1), the word BATTALION is inserted in proper position. Thus:

B E C J I   B T M T L   T W B P Q . . .
. . . . B   A T T A L   I O N . . . . .

With this as a start, the decipherment may proceed forward or backward with ease. Thus:

B E C J I   B T M T L   T W B P Q   A Y M N Q   H V N E T   W A A L C   . . .
E A C H B   A T T A L   I O N C O   M M A N D   E R W I L   L P L A C   . . .

c. The foregoing example is based upon the so-called Vigenère method of encipherment $(\Theta_{k/2}=\Theta_{1/1};\ \Theta_{p/1}=\Theta_{c/2})$. If in encipherment the plain-text letter is sought in the cipher component, its equivalent taken in the plain component $(\Theta_{k/2}=\Theta_{1/1};\ \Theta_{p/2}=\Theta_{c/1})$, the steps in solution are identical, except that the list of cipher equivalents of probable words must be modified accordingly. For instance, BATTALION will now be enciphered by the sequence . . . . ZTAHLXGZ.

---

[1] The student is cautioned that the characteristics noted apply only to the case where two identical components are used, with the base letter A.

d. If reversed standard cipher alphabets are used, the word BATTALION will be enciphered by the sequence.. BHATPDUB, which also presents idiomorphic characteristics leading to the easy recognition of the word.

e. All the foregoing phenomena are based upon standard alphabets, but when mixed cipher components are used and these have been reconstructed, similar observations may be recorded and the results employed in the solution of additional messages enciphered by the same components.

32. Example of solution by the probable-word method.—*a.* The solution of messages enciphered by unknown mixed components will now be discussed by example. When the primary components are unknown, the observations noted under the preceding subparagraphs are, of course, not applicable; nevertheless solution is not difficult. Given the following three cryptograms, all intercepted on the same day, and therefore suspected of being related:

Message I

```
H U F I I   O C Q J J   I V Z O Z   V P D G O   V V V K W
U E W H U   U Q H U M   R Z V Q R   U A K V D   N N E Z V
G J P G H   A Y J D R   U W N G R   Y S K B L   Q V U X N
P H D P R   S V K Z P   P P K G S   L L P R V   R B H A K
W U A V W   Y U E Z Q   X A P Q Y   G P S V S   F N R A K
C I F G Z   U V C C P   D K C W V   X T W F M   R F K B V
R O Q O J   D R U W N   G R Y S K   B L
```

Message II

```
J U F I I   O C Q J J   I V Z O Z   I B F E J   S U B R J
S P K T S   R Z V X T   W F M R F   Q H H F O   R F J P D
G O V V V   K W U H E   N D B D D   R H W U N   K C M P D
G O V Z S   E N D B D   D R H W U   N P P K P   E Q O Y
```

Message III

```
F J U H F   F K D E N   A L U P Z   K Q M V B   J W V P K
E U B D D   R H W U M   R H V G P   D N C U J   C D Z C Y
R H U J U   F Z P Q P   Y Q C Y H   O E Q Z V   X K C Q F
T V H N S   V C C E J   P E A M P   A P O E P   B H M V J
U N M H H   W K C V G   D S W J A   E Q Z B U   F F Y U E
Z Q X A P   Q Y G P A   R P Z V X   C F N R A   K C I F G
Z U V C C   P D K C O   G J W Z H   A P U F Z   F V H A V
X M H F F   K M Y H S   T B S K C   V R Q I J   Y C P Z H
U H C B M   T H O F H
```

b. (1) There are many repetitions, their intervals show no common factor, and a uniliteral frequency distribution does not appear to be monoalphabetic. Plain-text auto-keying is suspected. The simplest assumption to make at the start is that single-letter introductory keys are being used, with the normal Vigenère method of encipherment, and that the plain component is the normal sequence. Attempts to solve any of the messages on the assumption that the cipher component is also the normal sequence being unsuccessful, it is next assumed that the cipher component is a mixed sequence. The 13-letter repetition J D R U W N G R Y S K B L and the 10-letter repetition P D G O V V V K W U are studied intensively. If a single-letter introductory key is being used, then these repetitions involve 14-letter and 11-letter plain-text sequences or words; if the normal Vigenère method of encipherment is in effect $(\theta_{k/2}=\theta_{/i1};\ \theta_{p/1}=\theta_{c/2})$, then the base letter is A. If the latter is true then a good word which would fit the 13-letter repetition is:

Key	.	R	E	C	O	N	N	A	I	S	S	A	N	C	E
Plain text	R	E	C	O	N	N	A	I	S	S	A	N	C	E	.
Cipher		J	D	R	U	W	N	G	R	Y	S	K	B	L	.

and a good word which would fit the 10-letter repetition is:

Key	.	O	B	S	E	R	V	A	T	I	O	N
Plain text	O	B	S	E	R	V	A	T	I	O	N	.
Cipher		P	D	G	O	V	V	V	K	W	U	.

(2) Inserting, in a mixed component, the values given by these two assumptions yields the following:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	R		A		J		S	T	I	N	G	B	C			K	L					V	W		Y	
	E		D		U													O	P							

(3) It is a simple matter to combine these two partial cipher components into a single sequence, and the two components are as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	R	E	A	D	J	U	S	T	I	N	G	B	C	F	H	K	L	M	O	P	Q	V	W	X	Y	Z

(4) With the primary components at hand, solution of the messages is now an easy matter.

c. The foregoing example uses an unknown mixed cipher component sliding against what was first assumed (and later proved) to be the normal direct sequence. When both primary components are unknown mixed sequences but are identical, solution is more difficult, naturally, because the results of assuming values for repeated sequences cannot be proved and established so quickly as in the foregoing example. Nevertheless, the general method indicated, and the application of the principles of indirect symmetry will lead to solution, if there is a fair amount of text available for study. When an introductory key of several letters is used, repetitions are much reduced and the problem becomes still more difficult but by no means insurmountable. Space forbids a detailed treatment of the method of solving these cases but it is believed that the student is in a position to develop these methods and to experiment with them at his leisure.

33. Concluding remarks on the solution of auto-key systems.—*a.* The type of solution elucidated in the preceding paragraph is based upon the successful application of the probable-word method. But sometimes the latter method fails because the commonly expected words may not be present after all. Hence, other principles and methods may be useful. Some of these methods, useful in special cases, are almost mechanical in their nature. Extension of the basic principles involved may lead to rather far-reaching complexities. However, because these methods are applicable only to somewhat special situations, and because they are somewhat involved they will be omitted from the text proper and placed in Appendix 1. The student who is especially interested in these cases may consult that appendix at his leisure.

b. It is thought that sufficient attention has been devoted to the solution of both cipher-text and plain-text auto-key systems to have demonstrated to the student that these cryptographic methods have serious weaknesses which exclude them from practical usage in military cryptography. Besides being comparatively slow and subject to error, they are rather easily solvable, even when unknown cipher alphabets are employed.

c. In both systems there are characteristics which permit of identifying a cryptogram as belonging to this class of substitution. Both cases will show repetitions in the cipher text. In cipher-text auto-keying there will be far fewer repetitions than in the original plain text, especially when introductory keys of more than 1-letter in length are employed. In plain-text auto-keying there will be nearly as many repetitions in the cipher text as in the original plain text unless long introductory keys are used. In either system the repetitions will show no constancy as regards intervals between them, and a uniliteral frequency distribution will show such messages to be polyalphabetic in nature. Cipher-text auto-keying may be distinguished from plain-text auto-keying by the appearance of the frequency distribution of the second member of sets of two letters separated by the length of the introductory key (see par. 26*b*, *c*). In the case of cipher-text auto-keying these frequency distributions will be monoalphabetic in nature; in plain-text auto-keying such frequency distributions will not show monoalphabetic characteristics.

SECTION IX

METHODS OF LENGTHENING OR EXTENDING THE KEY

Paragraph
Preliminary remarks 34
Extended and nonrepeating keys; the so-called "running-key system" 35
Other systems employing lengthy keying sequences 36

34. Preliminary remarks.—In paragraph 1*b* of this text it was stated that two procedures suggest themselves for eliminating the weaknesses introduced by periodicity of the type produced by simple, repeating-key methods. The first of these, when studied, embraced some of the very simple methods of suppressing or destroying periodicity, by such devices as interrupting the key and using variable-length groupings of plain text. It was demonstrated that subterfuges of this simple nature are inadequate to eliminate the weaknesses referred to, and must be discarded in any system intended to afford real security. The other alternative suggested in paragraph 1*b* therefore remains now to be investigated, viz, that of lengthening the keys to a point where there would seem to be an insufficient amount of text to enable the cryptanalyst to solve the traffic. Attempts toward this end usually consist in extending the key to such a length that the enemy cryptanalysts will have only a very limited number of periods to work with. The key may, indeed, be lengthened to a point where it becomes as long as, or longer than, the text to be enciphered, so that the key is used only once.

35. Extended and nonrepeating keys.—*a*. It is obvious that one of the simplest methods of lengthening the key to a message is to use a long phrase or even a complete sentence, provided it is not too long to remember. In addition to the difficulties that would be encountered in practical military cryptography in selecting long mnemonic phrases and sentences which would have to be imparted to many clerks, there is the fact that the probable-word method of solution still remains as a powerful tool in the hands of enemy cryptanalysts. And if only a word or two of the key can be reconstructed as a result of a fortunate assumption, it is obvious that the enemy cryptanalysts could readily guess the entire key from a fragment thereof, since any long phrase or sentence which is selected because it can easily be remembered is likely to be well known to many people.

b. There are, however, more or less simple methods of employing a short mnemonic key in order to produce a much longer key. Basically, any method of transposition applied to a single alphabetic sequence repeated several times will yield a fairly long key, which, moreover, has the advantage of being unintelligible and thus approaching a random selection of letters. For example, a numerical key may be derived from a word or a short phrase; this numerical key may then be applied as a columnar-transposition key for a rectangle within which the normal alphabet has been repeated a previously agreed upon number of times in a normal (left to right) or prearranged manner. The letters when transcribed from the transposition rectangle then become the successive letters for enciphering the plain text, using any desired type of primary components. Or, if a single transposition is not thought to be sufficiently secure, a double transposition will yield a still more mixed up sequence of key letters. Other types of transposition may be employed for the purpose, including various kinds of geometric figures. Also, a non-transposition method of lengthening the keying sequence and at the same time introducing an irregularity, such as aperiodic interruption has already been described (see par. 18).

c. Another method of developing a long key from a short mnemonic one is that shown below. Given the keyword CHRISTMAS, a numerical sequence is first derived and then one writes down successive sections of this numerical key, these sections terminating with the successive numbers 1, 2, 3, . . . of the numerical key. Thus:

Mnemonic key....... C H R I S T M A S
Numerical key...... 2-3-6-4-7-9-5-1-8

```
                                  1|2|  3|      4|            5|    6|
Extended key....... C H R I S T M A|C|C H|C H R I|C H R I S T M|C H R|
                            7|                8|          9|
                    C H R I S|C H R I S T M A S|C H R I S T|
```

Thus the original key of only 9 letters is expanded to one of 45 letters (1+2+3+ . . . +9=45). The longer key is also an interrupted key of the type noted under paragraph 17, but if the message is long enough to require several repetitions of the expanded key the encipherment becomes periodic and can be handled by the usual methods employed in solving repeating-key ciphers. If the basic key is fairly long, so that the expanded key becomes a quite lengthy sequence, then the message or messages may be handled in the manner explained in paragraph 20.

d. Another method of producing a rather long sequence of digits for keying purposes from a single key number is to select a number whose reciprocal when converted by actual division into its equivalent decimal yields a long series of digits. For example the reciprocal of 49, or 1/49, yields a sequence of 42 digits beginning .02040815 . . . . Such a number, coupled with a key word like CHRISTMAS, could be used for interrupted keying, the successive cipher alphabets being used for enciphering as many letters as are indicated by the successive digits. In the case of the example cited, the first digit is 0; hence the C alphabet would not be used. The next digit is 2; the H alphabet would be used for enciphering the first and second letters. The third digit is again 0; the R alphabet would not be used. The fourth digit is 4; the I alphabet would be used for enciphering the third, fourth, fifth, and sixth letters, and so on.

36. Other systems employing lengthy keying sequences.—*a. The so-called "running-key" system.*—To be mentioned in connection with this subject of extensive or lengthy keys is the cipher system known as the running-key, continuous-key, or nonrepeating-key system, in which the key consists of a sequence of elements which never repeats no matter how long the message to be enciphered happens to be. The most common and most practical source of such a key is that in which the plain text of a previously agreed-upon book serves as the source for successive key letters for encipherment.[1] The solution of this type of cipher, an accomplishment which was once thought impossible, presents some interesting phases and will be considered shortly. At this point it is merely desired to indicate that according to the running-key system the key for an individual message may be as long as the message and never repeat; but if a large group of correspondents employ the same key sequence, it may happen that there will be several messages in the same key and they will all begin with the same initial key letter; or, there will be several which will "overlap" one another with respect to the key, that is, they begin at different initial points in the keying sequence but one message soon overtakes the other, so that from that point forward all subsequent letters in both messages are enciphered by the same sequence of key letters.

---

[1] Sec. IX, *Advanced Military Cryptography.* See also footnote 8, page 71 of this text.

b. The so-called progressive-alphabet system.—In the so-called progressive-alphabet system the basic principle is quite simple. Two or more primary elements are arranged or provided for according to a key which may be varied from time to time; the interaction of the primary elements results in making available for cryptographic purposes a set of cipher alphabets; all the latter are employed in a fixed sequence or progression; hence the designation progressive-alphabet system. If the number of alphabets available for such use is rather small, and if the text to be enciphered is much longer than the sequence of alphabets, then the system reduces to a periodic method. But if the number of alphabets is large, so that the sequence is not repeated, then of course, the cryptographic text will exhibit no periodic phenomena.

c. The series of cipher alphabets in such a system constitutes a keying sequence. Once set up, often the only remaining element in the key for a specific message is the starting point in the sequence, that is, the initial cipher alphabet employed in enciphering a given message. If this keying sequence must be employed by a large group of correspondents, and if all messages employ the same starting point in the keying sequence, obviously the cryptograms may simply be superimposed without any preliminary testing to ascertain proper points for superimposition. The student has already been shown how cases of this sort may be solved. However, if messages are enciphered with varying starting points, the matter of superimposing them properly takes on a different aspect. This will soon be treated in detail.

d. The respective cipher alphabets constituting the entire complement of alphabets may be employed in a simple progression, that is, consecutively from a preselected initial point; or, they may be employed according to other types of progression. For example, if the system comprises 100 alphabets one might use them in the sequence 1, 3, 5, 7, . . . ; or 1, 4, 7, 10, . . . ; or irregular types of skipping may be employed.

e. In addition to the foregoing, there are, of course, a great many mechanical methods of producing a long key, such as those employed in mechanical or electrical cipher machines. In most cases these methods depend upon the interaction of two or more short, primary keys which jointly produce a single, much longer, secondary or resultant key. (See par. 4.) Only brief reference can be made at this point in the cryptanalytic studies to cases of this kind. A detailed treatment of complex examples would require much time and space so that it will be reserved for subsequent texts.

f. Finally, there must be mentioned certain devices in which, as in encipherment by the auto-key method, the text itself serves to produce the variation in cipher equivalents, by controlling the selection of secondary alphabets, or by influencing or determining the sequence with which they will be employed. Naturally, in such cases the key is automatically extended to a point where it coincides in length with that of the text. An excellent example of such a device is that known as the Wheatstone, the solution of which will be described in its proper place.[2] Some writers classify and treat this method as well as auto-key methods as forms of the running-key system but the present author prefers to consider the latter as being radically different in principle from the former types, because in the true running-key system the key is wholly external to and independent of text being enciphered. This is hardly true of auto-key systems or of systems such as the Wheatstone mentioned herein.

---

[2] See Sec. XII, *Advanced Military Cryptography*.

SECTION X

GENERAL PRINCIPLES UNDERLYING SOLUTION OF SYSTEMS EMPLOYING LONG OR CONTINUOUS KEYS

Paragraph
Solution when the primary components are known sequences 37
Solution of a running-key cipher when an unknown but intelligible key sequence is used and the primary components are known 38
Solution of a progressive-alphabet cipher when the primary components are known 39
General solution for ciphers involving a long keying sequence of fixed length and composition 40

37. **Solution when the primary components are known sequences.**—*a*. As usual, the solution of cases involving long or continuous keys will be treated under two headings: First, when the primary components are known sequences; second, when these elements are wholly unknown or partially unknown.

b. Since the essential purpose in using long keys is to prevent the formation of repetitive cycles within the text, it is obvious that in the case of very long keying sequences the cryptanalyst is not going to be able to take the text and break it up into a number of small cycles which will permit the establishment of monoalphabetic frequency distributions that can readily be solved, an end which he can attain all the more readily if to begin with he knows the primary sequences. But, there nearly always remains the cryptanalyst's last resort: the probable-word method. Inasmuch as this method is applicable to most of these cases, even to that of the running-key system, which perhaps represents the furthest extension of the principle of long keying sequences, an example using a cryptogram of the latter type will be studied.

38. **Solution of a running-key cipher when an unknown but intelligible key sequence is used and the primary components are known.**—*a*. In paragraph 36*a* mention was made of the so-called running-key, continuous-key, or nonrepeating-key system, in which the plain text of a previously agreed-upon book serves as the source for successive key letters for encipherment. Since the running-key system is entirely aperiodic, and the cipher text can therefore not be arranged in superimposed short cycles, as in the case of the repeating-key system, it would appear on first consideration to be "indecipherable" without the key.[1] But if the student will bear in mind that one of the practical methods of solving a repeating-key cipher is that of the probable word,[2] he will immediately see that the latter method can also be applied in solving this type of nonrepeating-key system. The essence of the matter is this: The cryptanalyst may assume the presence of a probable word in the text of the message; if he knows the primary components involved, and if the assumed word actually exists in the message, he can locate it by checking against the key, *since the latter is intelligible text*. Or, he may assume the presence of a probable word or even of a phrase such as "to the," "of the," etc., in the key text and check his assumption against the text of the message. Once he has forced

---

[1] At one time, indeed, this view was current among certain cryptographers, who thought that the principle of factoring the intervals between repetitions in the case of the repeating-key cipher formed the basis for the only possible method of solving the latter type of system. Since, according to this erroneous idea, factoring cannot be applied in the case of the running-key system (using a book as the key), therefore solution was considered to be impossible. How far this idea is from the truth will presently be seen. In this same connection see also footnote 8, page 71.

[2] See *Military Cryptanalysis*, Part II, par. 25.

such an entering wedge into either the message or the key, he may build upon this foundation by extending his assumptions for text alternately in the key and in the message, thus gradually reconstructing both. For example, given a cryptogram containing the sequence . . . HVGGLOWBESLTR . . ., suppose he assumes the presence of the phrase THAT THE in the key text and finds a place in the plain text where this yields MMUNITI . Thus, using reversed standard cipher alphabets:

Assumed key text.................. . . . T H A T T H E . . .
Cipher text........................ . . . H V G G L O W B E S L T R . . .
Resultant plain text............... . . . M M U N I T I . . .

This suggests the word AMMUNITION. The ON in the cipher text then yields PR as the beginning of the word after THE in the key text. Thus:

Assumed key text.................. . . . T H A T T H E P R . . .
Cipher text........................ . . . H V G G L O W B E S L T R . . .
Resultant plain text............... . . . M M U N I T I O N . . .

PR must be followed by a vowel, with O the most likely candidate. He finds that O yields W in the plain text, which suggests the word WILL. The latter then yields OTEC in the key, making the latter read THAT THE PROTEC . . . . Thus:

Assumed key text.................. . . . T H A T T H E P R O T E C . . .
Cipher text........................ . . . H V G G L O W B E S L T R . . .
Resultant plain text............... . . . M M U N I T I O N W I L L . . .

This suggests the words PROTECTION, PROTECTIVE, PROTECTING, etc. Thus extending one text a few letters serves to "coerce" a few more letters out of the other, somewhat as in the case of two boys who are running approximately abreast in a race; as soon as one boy gets a bit ahead the spirit of competition causes the other to overtake and pass the first one; then the latter puts forth a little more effort, overtakes and passes the second boy. Thus the boys alternate in overtaking and passing each other until the race is run. The only point in which the simile fails is that while the boys usually run forward all the time, that is, in a single direction, the cryptanalyst is free to work in two directions—forward and backward from an internal point in the message. He may, in the case of the example cited above, continue his building-up process by adding A to the front of MMUNITI as well as ON to the rear. If he reaches the end of his resources on one end, there remains the other end for experimentation. He is certainly unlucky if both ends terminate in complete words both for the message and for the key, leaving him without a single clue to the next word in either, and forcing him to a more intensive use of his imagination, guided only by the context.

b. In the foregoing illustration the cryptanalyst is assumed to have only one message available for his experimentation. But if he has two or more messages which either begin at identical initial points with reference to the key, or overlap one another with respect to the key, the reconstruction process described above is, of course, much easier and is accomplished much more quickly. *For if the messages have been correctly superimposed with reference to the key text, the addition of one or two letters to the key yields suggestions for the assumption of words in several messages.* The latter lead to the addition of several letters to the key, and so on, in an ever-widening circle of ideas for further assumptions, since as the process continues the context affords more and more of a basis for the work.

c. Of course, if sufficient of the key text is reconstructed, the cryptanalyst might identify the book that is being used for the key, and if available, his subsequent labors are very much simplified.

d. All the foregoing is, however, dependent not only upon the use of an intelligible text as the keying text but also upon having a knowledge of the primary components or cipher alphabets employed in the encipherment. Even if the primary components are differently mixed sequences, so long as they are known sequences, the procedure is quite obvious in view of the foregoing explanation. The training the student has already had is believed sufficient to indicate to him the procedure he may follow in that solution, and no further details will here be given in respect to such cases. But what if the primary components are not known sequences? This contingency will be treated presently.

39. Solution of a progressive-alphabet cipher when the cipher alphabets are known.—*a.* Taking a very simple case, suppose the interacting elements referred to in paragraph 36*b* consist merely of two primary cipher components which slide against each other to produce a set of 26 secondary cipher alphabets, and suppose that the simplest type of progression is used, *viz*, the cipher alphabets are employed one after the other consecutively. Beginning at an initial juxtaposition, producing say, alphabet 1, the subsequent secondary alphabets are in the sequence 2, 3, . . . 26, 1, 2, 3, . . ., and so on. If a different initial juxtaposition is used, say alphabet 10 is the first one, the sequence is exactly the same as before, only beginning at a different point.

b. Suppose the two primary components are based upon the keyword HYDRAULIC. A message is to be enciphered, beginning with alphabet 1. Thus:

Plain component	H Y D R A U L I C B E F G J K M N O P Q S T V W X Z H Y D . . .
Cipher component	H Y D R A U L I C B E F G J K M N O P Q S T V W X Z

Letter No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Alphabet No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Plain text	E	N	E	M	Y	H	A	S	P	L	A	C	E	D	H	E	A	V	Y	I	N
Cipher text	E	O	G	P	U	U	E	Y	H	M	K	Q	V	M	K	Z	S	J	Q	H	E

Letter No.	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
Alphabet No.	22	23	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13
Plain text	T	E	R	D	I	C	T	I	O	N	F	I	R	E	U	P	O	N
Cipher text	N	L	H	H	L	C	V	B	S	S	N	J	E	P	K	D	D	D

Letter No.	40	41	42	43	44	45	46	47	48	49	50	51	52	53
Alphabet No.	14	15	16	17	18	19	20	21	22	23	24	25	26	1
Plain text	Z	A	N	E	S	V	I	L	L	E	R	O	A	D
Cipher text	G	P	U	H	F	K	H	H	Y	L	H	M	R	D

c. This method reduces to a periodic system involving 26 secondary cipher alphabets and the latter are used in simple progression. It is obvious therefore that the 1st, 27th, 53d, . . . letters are in the 1st alphabet; the 2d, 28th, 54th, . . . letters are in the 2d alphabet, and so on.

d. To solve such a cryptogram, knowing the two primary components, is hardly a problem at all. The only element lacking is a knowledge of the starting point. But this is not necessary, for merely by completing the plain-component sequences and examining the diagonals of the diagram, the plain text becomes evident. For example, given the following: HIDCT EHUXI. Completing the plain-component sequences initiated by the successive cipher letters, the plain text, ENEMY MACHI . . . is seen to come out in successive steps upward in Figure 10. Had the cipher component been shifted in the opposite direction in encipherment, the steps would have been downward instead of upward. If the sliding strips had been set up according to the sequence of cipher letters but on a diagonal, then, of course, the plain-text letters would have reappeared on one generatrix.

H	I	D	C	T	E	H	U	X	L
Y	C	R	B	V	F	Y	L	Z	I
D	B	A	E	W	G	D	I	H	C
R	E	U	F	X	J	R	C	Y	B
A	F	L	G	Z	K	A	B	D	E
U	G	I	J	H	M	U	E	R	F
L	J	C	K	Y	N	L	F	A	G
I	K	B	M	D	O	I	G	U	J
C	M	E	N	R	P	C	J	L	K
B	N	F	O	A	Q	B	K	I	M
E	O	G	P	U	S	E	M	C	N
F	P	J	Q	L	T	F	N	B	O
G	Q	K	S	I	V	G	O	E	P

FIGURE 10.

e. The student will understand what simple modifications in procedure would be required in case the two primary components were different mixed sequences. But what if the primary components are not known sequences? How does the cryptanalyst proceed in that case?

40. **General solution for ciphers involving a long-keying sequence of fixed length and composition.**—*a.* It is obvious, as stated at a previous point, that no matter how the keying sequence is derived, *if all the correspondents employ the same key, or if this key is used many times by a single office, and if it always begins at the same point, the various messages may simply be superimposed.* Thus, their respective 1st, 2d, 3rd, . . . letters will all fall within columns which have been enciphered by the 1st, 2d, 3rd, . . . key letters. If there is a sufficient number of messages, solution then becomes possible by frequency analysis of the successive columns—no matter how long the keying sequence may be, and regardless of whether the keying sequence constitutes intelligible text or is a purely random sequence of letters. This method of solution by superimposition has already been outlined in paragraph 20 and no further reference to it need here be made.

b. But now suppose that the keying sequence does not always begin at the same point for all messages. Suppose the several correspondents are able to select at will *any* point in the keying sequence as the point of departure in encipherment. Thus, such a keying sequence, if regarded as partaking of the nature of a circle, will afford as many possible starting points as there are letters or characters in that sequence. Now if there are no external indications or *indicators*[3] in the cryptograms pertaining to such a system, such as would afford enemy cryptanalysts direct and definite information with regard to the initial keying element for each cryptogram, then it would seem as though the superimposition of messages (to bring letters enciphered by the same cipher alphabets within the same columns) would be difficult or impossible, and therefore that attempts at solution are blocked at their very beginning. This, however, is not the end of the story. For suppose two of the messages have in common only one polygraph, say of 5 letters; these two messages may be juxtaposed so as to bring these repetitions into superimposition. Thus, the possession of this long polygraph in common serves to "tie" these two messages together or to "interlock" them. Then, suppose a shorter polygraph, say of 4 letters, is possessed in common by one of these two messages and a third message; this will serve to tie in the latter with the first two. Extension of this process, including the data from shorter repetitions of trigraphs and digraphs, will serve to assemble a whole set of such messages in proper superimposition. Therefore, the first step is to examine all the messages for repetitions.

---

[3] Indicators play an important rôle in practical cryptography. An indicator is a symbol (consisting of a letter, group of letters, a figure or a group of figures) which indicates the specific key used under the general cryptographic system, or it may indicate which one of a number of general systems has been used, or it may indicate both.

c. When such repetitions are found, and if there are plenty of them so that assumptions for probable words are easy to make, it is clear that the correct assumptions will enable the cryptanalyst to set up plain-cipher equivalencies which will make it possible to reconstruct the primary components. Depending upon the type used, the principles of direct or indirect symmetry of position will be very useful in this process.

d. But if it happens that there are no polygraphs by means of which two or more messages may be tied together and properly superimposed, the simple methods mentioned in subparagraphs *a–c* cannot here be applied. However, although the road toward a solution seems to be blocked rather effectively, there is a detour which presents rather interesting vistas. The latter are really of such importance in cryptanalysis as to warrant detailed treatment.

SECTION XI

THE "COINCIDENCE" OR "κ" TEST

Paragraph
The basic theory of the coincidence or κ (kappa) test........ 41
General procedure to be followed in making the κ test........ 42
Example of application of the κ test........ 43
Subsequent steps........ 44

41. The basic theory of the coincidence or κ (kappa) test.—*a.* In Appendix 2 of the preceding text[1] certain simple applications of the theory of probability were presented for the student's consideration, by way of pointing out to him the important role which certain phases of that branch of mathematics play in cryptanalysis. Reference was there made to the subject *of coincidences* and its significance in connection with the study of repetitions in cryptograms. In this section the matter will be pursued a few steps further.

b. In the appendix referred to, it was shown that the probability of monographic coincidence (1) in random text employing a 26-letter alphabet is .0385; (2) in English telegraphic plain text, .0667. These two parameters were represented by the symbols κ_r and κ_p, respectively. The important role which these values play in a certain cryptanalytic test will now be explained.

c. One of the most important techniques in cryptanalytics is that known as *applying the coincidence or "kappa test."* This test is useful for several cryptanalytic purposes and one of the most important of them is to ascertain when two or more sequences of letters are correctly superimposed. By the word "correct" in this case is merely meant that the sequences are so arranged relative to one another as to facilitate or make possible a solution. The test has for its theoretical basis the following circumstances:

(1) If any two rather lengthy sequences of characters are superimposed, it will be found, on examining both members of the successive pairs of letters brought into vertical juxtaposition, that *in a certain number of cases the two superimposed letters will coincide.*

(2) If both sequences of letters constitute random text (of a 26-letter alphabet), there will be about 38 or 39 such cases of coincidence per thousand pairs examined. This, of course, is because $\kappa_r = .0385$.

(3) If both sequences of letters constitute plain text, there will be about 66 or 67 such cases of coincidence per thousand pairs examined. This is because κ_p is .0667.

(4) If the superimposed sequences are wholly monoalphabetic encipherments of plain text by the same cipher alphabet, there will still be about 66 or 67 cases of coincidence in each 1,000 cases examined, because in monoalphabetic substitution there is a fixed or unvarying relation between plain-text letters and cipher letters so that for statistical purposes monoalphabetic cipher text behaves just the same as if it were normal plain text.

(5) Even if the two superimposed sequences are not monoalphabetically enciphered texts, but are polyalphabetic in character, there will still be about 66 or 67 cases of identity between superimposed letters per thousand cases examined, *provided the two sequences really belong to the same cryptographic system and are superimposed at the proper point with respect to the keying sequence.* The reasons for this will be set forth in the succeeding subparagraphs.

---

[1] *Military Cryptanalysis, Part II.* It is recommended that the student refresh his memory by reviewing that appendix.

(6) Consider the two messages below. They have been enciphered polyalphabetically by the same two primary components sliding against each other. The two messages use the same keying sequence, beginning at the same initial point in that sequence. Consequently, the two messages are identically enciphered, letter for letter, and the only differences between them are those occasioned by differences in plain text.

No. 1	Alphabets	16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7	12	6	
	Plain text	W	H	E	N	I	N	T	H	E	C	O	U	R	S	E	L	O	N	G	M	...
	Cipher	E	Q	N	B	T	F	Y	R	C	X	X	L	Q	J	N	Z	O	Y	A	W	...
No. 2	Alphabets	16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7	12	6	
	Plain text	T	H	E	G	E	N	E	R	A	L	A	B	S	O	L	U	T	E	L	Y	...
	Cipher	P	Q	N	T	U	F	B	W	D	J	L	Q	H	Y	Z	P	T	M	Q	I	...

Note, now, that (a) in every case in which two superimposed cipher letters are the same, the plain-text letters are identical and (b) in every case in which two superimposed cipher letters are different, the plain-text letters are different. In such a system, even though the cipher alphabet changes from letter to letter, the number of cases of identity or coincidence in the two members of a pair of superimposed cipher letters will still be about 66 or 67 per thousand cases examined, *because the two members of each pair of superimposed letters are in the same cipher alphabet and it has been seen in (4) that in monoalphabetic cipher text κ is the same as for plain text,*[2] *viz, .0667.* The two messages may here be said to be superimposed "correctly," that is, brought into proper juxtaposition with respect to the keying sequence.

(7) But now suppose the same two messages are superimposed "incorrectly," that is, they are no longer in proper juxtaposition with respect to the keying sequence. Thus:

No. 1	Alphabets	16	21	13	5	6	4	17	19	21	21	2	6	3	5	13	13	1	7	12		
	Plain text	W	H	E	N	I	N	T	H	E	C	O	U	R	S	E	L	O	N	G		...
	Cipher	E	Q	N	B	T	F	Y	R	C	X	X	L	Q	J	N	Z	O	Y	A		...
No. 2	Alphabets		16	21	13	5	6	4	17	19	21	21	2	6	3	6	13	13	1	7		
	Plain text		T	H	E	G	E	N	E	R	A	L	A	B	S	O	L	U	T	E		...
	Cipher		P	Q	N	T	U	F	B	W	D	J	L	Q	H	Y	Z	P	T	M		...

It is evident that the two members of every pair of superimposed letters are no longer in the same cipher alphabet, and therefore, if two superimposed cipher letters *are* identical this is merely an "accident," for now there is no basic or general cause for the similarity, such as is true in the case of a correct superimposition. The similarity, if present, is, as already stated, due to chance and the number of such cases of similarity should be about the same as though the two cipher letters were drawn at random from random text, in which $\kappa_r = .0385$. It is no longer true that (a) in every case in which two superimposed cipher letters are the same, the plain-text letters are identical, or (b) in every case in which two superimposed cipher letters are different, the plain-text letters are different. Note, for example, that the superimposed T_c's represent two different plain-text letters and that the S_p of the word COURSE in the first message gives J_c while the S of the word ABSOLUTELY in the second message gives H_c. Thus, it becomes clear that in an incorrect superimposition two different plain-text letters enciphered by two different alphabets may "by chance" produce identical cipher letters, which on superimposition yield a coincidence having no external indications as to dissimilarity in plain-text equivalents. Hence, if there are no other factors which enter into the matter and which might operate to distort the results to be expected from the operation of the basic factor, the expected number of cases of identical cipher letters brought together by an incorrect superimposition will be determined by the value $\kappa_r = .0385$.

---

[2] The fact that in this case each monoalphabet contains but two letters does not affect the theoretical value of κ; and whether the actual number of coincidences agrees closely with the expected number based upon $\kappa = .0667$ depends upon the lengths of the two superimposed sequences.

(8) But now note also that in the foregoing incorrect superimposition there are two Z_c's and that they represent the same plain-text letter L. This is occasioned by the fact that the plain-text messages happened to have L's in just those two places and that the cipher alphabet happened to be the same both times. Hence, it becomes clear that the same cipher alphabet brought into play twice may "by chance" happen to encipher the same plain-text letter both times, thus producing identical cipher letters. In some systems this source of identity in superimposed cipher letters is of little importance, in other systems, it may materially affect the actual number of coincidences. For instance, if a system is such that it produces a long secondary keying cycle composed of repetitions of short primary keying cycles, an incorrect superimposition of two cryptograms may bring into juxtaposition many of these short cycles, with the result that the actual number of cases of identical superimposed cipher letters is much greater than the expected number based upon $\kappa_r = .0385$. Thus, this source for the production of identical cipher letters in an incorrect superimposition operates to increase the number of cases to be expected from the fundamental constant $\kappa_r = .0385$.

(9) In some systems, where nonrelated cipher alphabets are employed, it may happen that two identical plain-text letters may be enciphered by two different cipher alphabets which, "by chance," have the same equivalent for the plain-text letter concerned. This is, however, a function of the particular cryptographic system and can be taken into account when the nature of the system is known.

(10) In general, then, it may be said that in the case of a correct superimposition the probability of identity or coincidence in superimposed cipher letters is .0667; in the case of an incorrect superimposition, the probability is at least .0385 and may be somewhat greater, depending upon special circumstances. The foregoing situation and facts make possible what has been referred to as the "coincidence test." Since this test uses the constant κ, it is also called the "kappa test."

d. The way in which the coincidence test may be applied will now be explained. The statement that $\kappa_p = .0667$ means that in 1,000 cases where two letters are drawn at random from a large volume of plain text, there will be about 66 or 67 cases in which the two letters coincide, that is, are identical. Nothing is specified as to what the two letters shall be; they may be two Z's or they may be two E's. This constant, .0667, really denotes a percentage: If many *comparisons* of single letters are made, the letters being drawn *at random* from among those constituting a large volume of plain text, 6.67 percent of these comparisons made will yield coincidences. So, if 2,000 such comparisons are made, the theory indicates that there should be about $.0667 \times 2{,}000 = 133$ coincidences; if there is sufficient text to permit of making 20,000 comparisons, there should be about 1,334 coincidences, and so on.

e. Another way of handling the matter is to find the ratio of the observed number of coincidences to the total number of cases in which the event in question might possibly occur, i. e., the total number of comparisons of superimposed letters. When this ratio is closer to .0667 than it is to .0385 the correct superimposition has been ascertained. This is true because in the case of a correct superimposition both members of each pair of superimposed letters actually belong to the same monoalphabet and therefore the probability of their coinciding is .0667; whereas in the case of an incorrect superimposition the members of each pair of superimposed letters belong, as a general rule, to different monoalphabets[3], and therefore the probability of their coinciding is nearer .0385 than .0667.

f. From the foregoing, it becomes clear that the kappa test involves ascertaining the total number of comparisons that can be made in a given case, as well as ascertaining the actual number of coincidences in the case under consideration. When only two messages are superimposed, this is easy: The total number of comparisons that can be made is the same as the number of superimposed pairs of letters. But when more than two messages are superimposed in a *superimposition diagram* it is necessary to make a simple calculation, based upon the fact that n letters yield $\frac{n(n-1)}{2}$ pairs or comparisons, where n is the number of letters in the column.[4] For example, in the case of a column of 3 letters, there are $\frac{3\times 2}{2}=3$ comparisons. This can be checked by noting that the 1st letter in the column may be compared with the 2d, the 2d with the 3d, and the 1st with the 3d, making 3 comparisons in all. The number of comparisons per column times the number of columns in the superimposition diagram of letters gives the total number of comparisons. The extension of this reasoning to the case where a superimposition diagram has columns of various lengths is quite obvious: one merely adds together the number of comparisons for columns of different lengths to obtain a grand total. For convenience, the following brief table is given:

Number of letters in column	Number of comparisons	Number of letters in column	Number of comparisons	Number of letters in column	Number of comparisons
2	1	11	55	21	210
3	3	12	66	22	231
4	6	13	78	23	253
5	10	14	91	24	276
6	15	15	105	25	300
7	21	16	120	26	325
8	28	17	136	27	351
9	36	18	153	28	378
10	45	19	171	29	406
		20	190	30	435

g. In ascertaining the number of coincidences in the case of a column containing several letters, it is again necessary to use the formula $\frac{n(n-1)}{2}$, only in this case n is the number of identical letters in the column. The reasoning, of course, is the same as before. The total number of coincidences is the sum of the number of coincidences for each case of identity. For example, in the column shown at the side, containing 10 letters, there are 3 B's, 2 C's, 4 K's, and 1 Z. The 3 B's yield 3 coincidences, the 2 C's yield 1 coincidence, and the 4 K's yield 6 coincidences. The sum of 3+1+6 makes a total of 10 coincidences in 45 comparisons.

---

[3] The qualifying phrase "as a general rule" is intended to cover any distortion in results occasioned by the presence of an unusual number of those cases of coincidence described under subpar. *c* (8) and (9).

[4] This has already been encountered (footnote 3, Appendix 2, *Military Cryptanalysis, Part II*). It is merely a special case under the general formula for ascertaining the number of combinations that may be made of n different things taken r at a time, which is ${}_nC_r=\frac{n!}{r!(n-r)!}$. In studying coincidences by the method indicated, since only two letters are compared at a time, r is always 2; hence the expression $\frac{n!}{r!(n-r)!}$, which is the same as $\frac{n(n-1)(n-2)!}{2(n-2)!}$, becomes by cancellation of $(n-2)!$, reduced to $\frac{n(n-1)}{2}$.

C
K
B
K
Z
K
C
B
B
K

42. General procedure to be followed in making the κ test.—*a*. The steps in applying the foregoing principles to an actual case will now be described. Suppose several messages enciphered by the same keying sequence but each beginning at a different point in that sequence are to be solved. The indicated method of solution is that of superimposition, the problem being to determine just where the respective messages are to be superimposed so that the cipher text within the respective columns formed by the superimposed messages will be monoalphabetic. From what has been indicated above, it will be understood that the various messages may be shifted relative to one another to many different points of superimposition, there being but one correct superimposition for each message with respect to all the others. First, all the messages are numbered according to their lengths, the longest being assigned the number 1. Commencing with messages 1 and 2, and keeping number 1 in a fixed position, message 2 is placed under it so that the initial letters of the two messages coincide. Then the two letters forming the successive pairs of superimposed letters are examined and the total number of cases in which the superimposed letters are identical is noted, this giving the observed number of coincidences. Next, the total number of superimposed pairs is ascertained, and the latter is multiplied by .0667 to find the expected number of coincidences. If the observed number of coincidences is considerably below the expected number, or if the ratio of the observed number of coincidences to the total number of comparisons is nearer .0385 than .0667, the superimposition is incorrect and message 2 is shifted to the next superimposition, that is, so that its first letter is under the second of message 1. Again the observed number of coincidences is ascertained and is compared with the expected number. Thus, by shifting message 2 one space at a time (to the right or left relative to message 1) the coincidence test finally should indicate the proper relative positions of the two messages. When the correct point of superimposition is reached the cryptanalyst is rarely left in doubt, for the results are sometimes quite startling. After messages 1 and 2 have been properly superimposed, message 3 is tested first against messages 1 and 2 separately, and then against the same two messages combined at their correct superimposition.[5] Thus message 3 is shifted a step each time until its correct position with respect to messages 1 and 2 has been found. Then message 4 is taken and its proper point of superimposition with respect to messages 1, 2, and 3 is ascertained. The process is continued in this manner until the correct points of superimposition for all the messages have been found. It is obvious that as messages are added to the superimposition diagram, the determination of correct points of superimposition for subsequent messages becomes progressively more certain and therefore easier.

b. In the foregoing procedure it is noted that there is necessity for repeated displacement of one message against another or other messages. Therefore, it is advisable to transcribe the messages on long strips of cross-section paper, joining sections accurately if several such strips are necessary to accommodate a long message. Thus, a message once so transcribed can be shifted to various points of superimposition relative to another such message, without repeatedly rewriting the messages.

c. Machinery for automatically comparing letters in applying the coincidence test has been devised. Such machines greatly facilitate and speed up the procedure.

---

[5] At first thought the student might wonder why it is advisable or necessary to test message 3 against message 1 and message 2 separately before testing it against the combination of messages 1 and 2. The first two tests, it seems to him, might be omitted and time saved thereby. The matter will be explained in par. 43*f* (3).

43. Example of application of the κ test.—*a.* With the foregoing in mind, a practical example will now be given. The following messages, assumed to be the first 4 of a series of 30 messages, supposedly enciphered by a long keying sequence, but each message commencing at a different point in that sequence, are to be arranged so as to bring them into correct superimposition:

MESSAGE 1

```
PGLPN  HUFRK  SAUQQ  AQYUO  ZAKGA  EOQCN
PRKOV  HYEIU  YNBON  NFDMW  ZLUKQ  AQAHZ
MGCDS  LEAGC  JPIVJ  WVAUD  BAHMI  HKORM
LTFYZ  LGSOG  K
```

MESSAGE 2

```
CWHPK  KXFLU  MKURY  XCOPH  WNJUW  KWIHL
OKZTL  AWRDF  GDDEZ  DLBOT  FUZNA  SRHHJ
NGUZK  PRCDK  YOOBV  DDXCD  OGRGI  RMICN
HSGGO  PYAOY  X
```

MESSAGE 3

```
WFWTD  NHTGM  RAAZG  PJDSQ  AUPFR  OXJRO
HRZWC  ZSRTE  EEVPX  OATDQ  LDOQZ  HAWNX
THDXL  HYIGK  VYZWX  BKOQO  AZQND  TNALT
CNYEH  TSCT
```

MESSAGE 4

```
TULDH  NQEZZ  UTYGD  UEDUP  SDLIO  LNNBO
NYLQQ  VQGCD  UTUBQ  XSOSK  NOXUV  KCYJX
CNJKS  ANGUI  FTOWO  MSNBQ  DBAIV  IKNWG
VSHIE  P
```

b. Superimposing[6] messages 1 and 2, beginning with their 1st letters,

5 10 15 20 25 30 35

No. 1...... P G L $\underline{\text{P}}$ N H U $\underline{\text{F}}$ R K S A $\underline{\text{U}}$ Q Q A Q Y U O Z A K G A E O Q C N P R K O V

No. 2...... C W H $\underline{\text{P}}$ K K X $\underline{\text{F}}$ L U M K $\underline{\text{U}}$ R Y X C O P H W N J U W K W I H L O K Z T L

40 45 50 55 60 65 70

No. 1...... H Y E I U Y N B O N N F D M W Z L U K Q A Q A $\underline{\text{H}}$ Z M $\underline{\text{G}}$ C D S L E A G C

No. 2...... A W R D F G D D E Z D L B O T F U Z N A S R H $\underline{\text{H}}$ J N $\underline{\text{G}}$ U Z K P R C D K

75 80 85 90 95 100

No. 1...... J P I V J W V A U $\underline{\text{D}}$ B A H M $\underline{\text{I}}$ H K O R M L T F Y Z L G S $\underline{\text{O}}$ G K

No. 2...... Y O O B V D D X C $\underline{\text{D}}$ O G R G $\underline{\text{I}}$ R M I C N H S G G O P Y A $\underline{\text{O}}$ Y X

the number of coincidences is found to be 8. Since the total number of comparisons is 101, the expected number, if the superimposition were correct, should be $101 \times .0667 = 6.7367$, or about 7 coincidences. The fact that the observed number of coincidences matches and is even greater than the expected number on the very first trial creates an element of suspicion: such good fortune is rarely the lot of the practical cryptanalyst. It is very unwise to stop at the first trial, *even if the results are favorable,* for this close agreement between theoretical and actual numbers of coincidences might just be "one of those accidents." Therefore message 2 is shifted one space to the right, placing its 1st letter beneath the 2d letter of message 1. Again the number of coincidences is noted and this time it is found to be only 4. The total number of comparisons is now 100; the expected number is still about 7. Here the observed number of coincidences is considerably less than the expected number, and when the relatively small number of comparisons is borne in mind, the discrepancy between the theoretical and actual results is all the more striking. The hasty cryptanalyst might therefore jump to the conclusion that the 1st superimposition is actually the correct one. But only two trials have been made thus far and a few more are still advisable, for in this scheme of superimposing a series of messages it is absolutely essential that the very first superimpositions rest upon a perfectly sound foundation—otherwise subsequent work will be very difficult, if not entirely fruitless. Additional trials will therefore be made.

---

[6] The student will have to imagine the messages written out as continuous sequences on cross-section paper.

c. Message 2 is shifted one more space to the right and the number of coincidences is now found to be only 3. Once again message 2 is shifted, to the position shown below, and the observed number of coincidences jumps suddenly to 9.

```
                     5        10        15        20        25        30        35
No. 1....... P G L P N H U F R K S A U Q Q A Q Y U O Z A K G A E O Q C N P R K O V
No. 2.......       C W H P K K X F L U M K U R Y X C O P H W N J U W K W I H L O K

                    40        45        50        55        60        65        70
No. 1....... H Y E I U Y N B O N N F D M W Z L U K Q A Q A H Z M G C D S L E A G C
No. 2....... Z T L A W R D F G D D E Z D L B O T F U Z N A S R H H J N G U Z K P R

                    75        80        85        90        95       100
No. 1....... J P I V J W V A U D B A H M I H K O R M L T F Y Z L G S O G K
No. 2....... C D K Y O O B V D D X C D O G R G I R M I C N H S G G O P Y A O Y X
```

The total number of comparisons is now 98, so that the expected number of coincidences is 98 × .0667 = 6.5366, or still about 7. The 2d and 3d superimpositions are definitely incorrect; as to the 1st and 4th, the latter gives almost 30 percent more coincidences than the former. Again considering the relatively small number of comparisons, this 30 percent difference in favor of the 4th superimposition as against the 1st is important. Further detailed explanation is unnecessary, and the student may now be told that it happens that the 4th superimposition is really correct; if the messages were longer, all doubt would be dispelled. The relatively large number of coincidences found at the 1st superimposition is purely accidental in this case.

d. The phenomenon noted above, wherein the observed number of coincidences shows a *sudden* increase in moving from an incorrect to a correct superimposition is not at all unusual, nor should it be unexpected, because there is only *one* correct superimposition, while *all other* superimpositions are entirely incorrect. In other words, a superimposition is either 100 percent correct or 100 percent wrong—and there are no gradations between these two extremes. Theoretically, therefore, the difference between the correct superimposition and any one of the many incorrect superimpositions should be very marked, since it follows from what has been noted above, that one cannot expect that the discrepancy between the actual and the theoretical number of coincidences should get smaller and smaller as one approaches closer and closer to the correct superimposition.[7] For if letters belonging to the same cipher alphabet are regarded as being members of the same family, so to speak, then the two letters forming the successive pairs of letters brought into superimposition by an incorrect placement of one message relative to another are total strangers to each other, brought together by pure chance. This happens time and again, as one message is slid against the other—until the correct superimposition is reached, whereupon in *every* case the two superimposed letters belong to the same family. There may be many different families (cipher alphabets) but the fact that in every case two members of the same family are present causes the marked jump in number of coincidences.

---

[7] The importance of this remark will be appreciated when the student comes to study longer examples, in which statistical expectations have a better opportunity to materialize.

e. In shifting one message against another, the cryptanalyst may move to the right constantly, or he may move to the left constantly, or he may move alternately to the left and right from a selected initial point. Perhaps the latter is the best plan.

f. (1) Having properly superimposed messages 1 and 2, message 3 is next to be studied. Now it is of course possible to test the latter message against the combination of the former, without further ado. That is, ascertaining merely the total number of coincidences given by the superimposition of the 3 messages might be thought sufficient. But for reasons which will soon become apparent it is better, even though much more work is involved, first to test message 3 against message 1 alone and against message 2 alone. This will really not involve much additional work after all, since the two tests can be conducted simultaneously, because the proper superimposition of messages 1 and 2 is already known. If the tests against messages 1 and 2 separately at a given superimposition give good results, then message 3 can be tested, at that superimposition, against messages 1 and 2 combined. That is, all 3 messages are tested as a single set. Since, according to the scheme outlined, a set of three closely related tests is involved, one might as well systematize the work so as to save time and effort, if possible. With this in view a diagram such as that shown in Figure 11*a* is made and in it the coincidences are recorded in the appropriate cells, to show separately the coincidences between messages 1 and 2, 1 and 3, 2 and 3, for each superimposition tested. The number of tallies in the cell 1–2 is the same at the beginning of all the tests; it has already been found to be 9. Therefore, 9 tallies are inserted in cell 1–2 to begin with. A column which shows identical letters in messages 1 and 3 yields a single tally for cell 1–3; a column which shows identical letters in messages 2 and 3 yields a single tally for cell 2–3. Only when a superimposition yields 3 identical letters in a column, is a tally to be recorded simultaneously in cells 1–3 and 2–3, since the presence of 3 identical letters in the column yields 3 coincidences.

	1	2	3
1	×	𝍸 ////	///
2	×	×	///
3	×	×	×

FIGURE 11*a*.

(2) Let message 3 be placed beneath messages 1 and 2 combined, so that the 1st letter of message 3 falls under the 1st letter of message 1. (It is advisable to fasten the latter in place so that they cannot easily be disturbed.) Thus:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	P	G	L	P	N	H	U	F	R	K	S	A	U	Q	Q	A	Q	Y	U	O	Z	A	K	G	A	E	O
2				C	W	H	P	K	K	X	F	L	U	M	K	U	R	Y	X	C	O	P	H	W	N	J	U
3	W	F	W	T	D	N	H	T	G	M	R	A	A	Z	G	P	J	D	S	Q	A	U	P	F	R	O	X

	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
1	Q	C	N	P	R	K	O	V	H	Y	E	I	U	Y	N	B	O	N	N	F	D	M	W	Z	L	U	K
2	W	K	W	I	H	L	O	K	Z	T	L	A	W	R	D	F	G	D	D	E	Z	D	L	V	O	T	F
3	J	R	O	H	R	Z	W	C	Z	S	R	T	E	E	E	V	P	X	O	A	T	D	Q	L	D	O	Q

	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
1	Q	A	Q	A	H	Z	M	G	C	D	S	L	E	A	G	C	J	P	I	V	J	W	V	A	U	D	B
2	U	Z	N	A	S	R	H	H	J	N	G	U	Z	K	P	R	C	D	K	Y	O	O	B	V	D	D	X
3	Z	H	A	W	N	X	T	H	D	X	L	H	Y	I	G	K	V	Y	Z	W	X	B	K	O	Q	O	A

	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104
1	A	H	M	I	H	K	O	R	M	L	T	F	Y	Z	L	G	S	O	G	K			
2	C	D	O	G	R	G	I	R	M	I	C	N	H	S	G	G	O	P	Y	A	O	Y	X
3	Z	Q	N	D	T	N	A	L	T	C	N	Y	E	H	T	S	C	T					

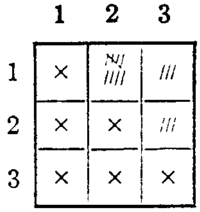


FIGURE 11b.

The successive columns are now examined and the coincidences are recorded, remembering that only coincidences between messages 1 and 3, and between messages 2 and 3 are now to be tabulated in the diagram. The results for this first test are shown in Figure 11b. This superimposition yields but 3 coincidences between messages 1 and 3, and the same number between messages 2 and 3. The total numbers of comparisons are then noted and the following table is drawn up:

Combination	Total number of comparisons	Number of coincidences		Discrepancy
		Expected	Observed	
				Percent
Messages 1 and 3	99	About 7	3	−57
Messages 2 and 3	96	About 6	3	−50
Messages 1, 2, and 3	293	About 20	15	−21

(3) The reason for the separate tabulation of coincidences between messages 1 and 3, 2 and 3. and 1. 2. and 3 should now be apparent. Whereas the observed number of coincidences is 57 percent below the expected number of coincidences in the case of messages 1 and 3 alone, and 50 percent below in the case of messages 2 and 3 alone, the discrepancy between the expected and observed numbers is not quite so marked (−21 percent) when all three messages are considered together, because the relatively high number of coincidences between messages 1 and 2, which are correctly superimposed, serves to counterbalance the low numbers of coincidences between 1 and 3, and 2 and 3. *Thus, a correct superimposition for one of the three combinations may yield such good results as to mask the bad results for the other two combinations.*

(4) Message 3 is then shifted one space to the right, and the same procedure is followed as before. The results are shown below:

```
                    5         10        15        20        25        30        35
No. 1...... P G L P N H U F R K S A U Q Q A Q Y U O Z A K G A E O Q C N P R K O V
No. 2......       C W H P K K X F L U M K U R Y X C O P H W N J U W K W I H L O K
No. 3......   W F W T D N H T G M R A A Z G P J D S Q A U P F R O X J R O H R Z W

                    40        45        50        55        60        65        70
No. 1...... H Y E I U Y N B O N N F D M W Z L U K Q A Q A H Z M G C D S L E A G C
No. 2...... Z T L A W R D F G D D E Z D L B O T F U Z N A S R H H J N G U Z K P R
No. 3...... C Z S R T E E E V P X O A T D Q L D O Q Z H A W N X T H D X L H Y I G

                    75        80        85        90        95          101
No. 1...... J P I V J W V A U D B A H M I H K O R M L T F Y Z L G S O G K
No. 2...... C D K Y O O B V D D X C D O G R G I R M I C N H S G G O P Y A O Y X
No. 3...... K V Y Z W X B K O Q O A Z Q N D T N A L T C N Y E H T S C T
```

	1	2	3
1	×	×	×
2	×	×	×
3	×	×	×

FIGURE 11c.

Combination	Total number of comparisons	Number of coincidences		Discrepancy
		Expected	Observed	
				Percent
Messages 1 and 3	99	About 7	10	+43
Messages 2 and 3	97	About 6	6	0
Messages 1, 2, and 3	294	About 20	25	+25

Note how well the observed and expected numbers of coincidences agree in all three combinations. Indeed, the results of this test are so good that the cryptanalyst might well hesitate to make any more tests.

(5) Having ascertained the relative positions of 3 messages, the fourth message is now studied. Here are the results for the correct superimposition.

					5					10					15					20					25					30					35
No. 1.......	P	G	L	P	N	$\underline{H}$	U	F	R	K	S	A	$\underline{U}$	Q	Q	A	Q	$\underline{Y}$	U	O	Z	$\underline{A}$	K	G	A	E	$\underline{O}$	Q	C	$\underline{N}$	P	R	K	$\underline{O}$	V
No. 2.......				C	W	$\underline{H}$	P	K	K	X	F	L	$\underline{U}$	M	K	U	R	$\underline{Y}$	X	C	O	$\underline{P}$	H	W	N	J	$\underline{U}$	W	K	$\underline{W}$	I	$\underline{H}$	L	$\underline{O}$	K
No. 3.......		W	F	W	T	$\underline{D}$	N	H	T	G	M	R	A	A	Z	$\underline{G}$	P	J	D	S	Q	$\underline{A}$	U	P	F	R	$\underline{O}$	X	J	R	O	$\underline{H}$	R	Z	W
No. 4.......			T	U	L	$\underline{D}$	H	N	Q	E	Z	Z	$\underline{U}$	T	Y	$\underline{G}$	D	U	E	D	U	$\underline{P}$	S	D	L	I	$\underline{O}$	L	N	$\underline{N}$	B	O	N	Y	L

					40					45					50					55					60					65					70
No. 1.......	H	Y	E	I	U	Y	N	B	O	N	N	F	D	M	W	Z	$\underline{L}$	U	K	$\underline{Q}$	A	Q	$\underline{A}$	H	Z	M	G	$\underline{C}$	$\underline{D}$	S	$\underline{L}$	E	$\underline{A}$	G	C
No. 2.......	Z	T	L	A	W	R	$\underline{D}$	F	G	D	D	E	Z	D	L	B	O	T	F	U	$\underline{Z}$	N	$\underline{A}$	S	R	H	H	J	$\underline{N}$	G	U	Z	K	P	R
No. 3.......	C	Z	S	R	T	E	E	E	V	P	X	O	A	T	D	Q	$\underline{L}$	D	$\underline{O}$	$\underline{Q}$	$\underline{Z}$	H	$\underline{A}$	W	N	X	T	H	$\underline{D}$	X	$\underline{L}$	H	Y	I	$\underline{G}$
No. 4.......	Q	Q	V	Q	G	C	$\underline{D}$	U	T	U	B	Q	X	S	O	S	K	N	$\underline{O}$	X	U	V	K	C	Y	J	X	$\underline{C}$	$\underline{N}$	J	K	S	$\underline{A}$	N	$\underline{G}$

					75					80					85					90					95						101						
No. 1.......	J	P	I	V	J	$\underline{W}$	V	A	U	$\underline{D}$	$\underline{B}$	$\underline{A}$	H	M	I	H	K	O	$\underline{R}$	$\underline{M}$	L	T	F	$\underline{Y}$	Z	L	$\underline{G}$	$\underline{S}$	O	G	K						
No. 2.......	C	D	K	Y	$\underline{O}$	O	$\underline{B}$	V	D	$\underline{D}$	X	C	$\underline{D}$	O	G	R	G	$\underline{I}$	$\underline{R}$	$\underline{M}$	I	$\underline{C}$	$\underline{N}$	H	S	G	$\underline{G}$	O	P	Y	A	O	Y	X			
No. 3.......	K	V	Y	Z	W	X	$\underline{B}$	K	O	Q	O	$\underline{A}$	Z	Q	N	D	T	N	A	L	T	$\underline{C}$	$\underline{N}$	$\underline{Y}$	E	H	T	$\underline{S}$	C	T							
No. 4.......	U	I	F	T	$\underline{O}$	$\underline{W}$	O	M	S	N	$\underline{B}$	Q	$\underline{D}$	B	A	I	V	$\underline{I}$	K	N	W	G	V	S	H	I	E	P									

	1	2	3	4
1	×	𝍸 ////	𝍸 𝍸	𝍸 //
2	×	×	𝍸 /	𝍸 /
3	×	×	×	𝍸
4	×	×	×	×

FIGURE 11d

Combination	Total number of comparisons	Number of coincidences		Discrepancy
		Expected	Observed	
				Percent
Messages 1 and 4	96	About 6	7	+16
Messages 2 and 4	95	About 6	7	+16
Messages 3 and 4	96	About 6	5	−16
Messages 1, 2, 3, and 4	581	About 39	44	+10

The results for an incorrect superimposition (1st letter of message 4 under 4th letter of message 1) are also shown for comparison:

					5					10					15					20					25					30					35
No. 1	P	G	L	P	N	$\underline{H}$	U	F	R	K	S	A	$\underline{U}$	Q	Q	A	Q	$\underline{Y}$	$\underline{U}$	O	Z	$\underline{A}$	K	G	A	E	$\underline{O}$	Q	C	$\underline{N}$	P	R	K	$\underline{O}$	V
No. 2				C	W	$\underline{H}$	P	K	K	X	F	L	$\underline{U}$	M	K	U	R	$\underline{Y}$	X	C	O	P	H	W	N	J	U	W	K	W	I	$\underline{H}$	L	$\underline{O}$	K
No. 3		W	F	W	T	D	N	$\underline{H}$	T	G	M	R	A	A	Z	G	P	J	D	S	Q	$\underline{A}$	U	P	F	R	$\underline{O}$	X	J	R	O	$\underline{H}$	R	Z	W
No. 4				T	U	L	D	$\underline{H}$	N	Q	E	Z	Z	U	T	Y	G	D	$\underline{U}$	E	D	U	P	S	D	L	I	O	L	$\underline{N}$	N	B	O	N	Y

					40					45					50					55					60					65					70
No. 1	H	Y	E	I	U	Y	N	B	O	N	N	F	D	M	W	Z	$\underline{L}$	U	K	$\underline{Q}$	A	Q	$\underline{A}$	H	Z	M	G	C	$\underline{D}$	S	$\underline{L}$	E	A	G	C
No. 2	Z	T	L	A	W	R	D	F	G	D	D	E	Z	D	L	B	O	T	F	U	$\underline{Z}$	N	$\underline{A}$	S	R	H	H	J	N	G	U	Z	K	P	R
No. 3	C	Z	S	R	T	E	E	E	V	P	X	O	A	T	D	Q	$\underline{L}$	D	O	$\underline{Q}$	$\underline{Z}$	H	$\underline{A}$	W	N	X	T	H	$\underline{D}$	X	$\underline{L}$	H	Y	I	G
No. 4	L	Q	Q	V	Q	G	C	D	U	T	U	B	Q	X	S	O	S	K	N	O	X	U	V	K	C	Y	J	X	C	N	J	K	S	A	N

					75					80					85					90					95						101			
No. 1	J	P	$\underline{I}$	V	J	W	V	A	U	$\underline{D}$	B	$\underline{A}$	H	M	I	H	K	O	$\underline{R}$	$\underline{M}$	L	T	F	Y	Z	L	$\underline{G}$	$\underline{S}$	O	G	K			
No. 2	C	D	K	Y	O	$\underline{O}$	$\underline{B}$	V	D	$\underline{D}$	X	C	D	O	G	R	G	I	$\underline{R}$	$\underline{M}$	I	$\underline{C}$	$\underline{N}$	H	$\underline{S}$	G	$\underline{G}$	O	$\underline{P}$	Y	A	O	Y	X
No. 3	K	V	Y	Z	W	X	$\underline{B}$	K	O	Q	O	$\underline{A}$	Z	Q	N	D	T	N	A	L	T	$\underline{C}$	$\underline{N}$	Y	E	$\underline{H}$	T	$\underline{S}$	C	T				
No. 4	G	U	$\underline{I}$	F	T	$\underline{O}$	W	O	M	S	N	B	Q	D	B	A	I	V	I	K	N	W	G	V	$\underline{S}$	$\underline{H}$	I	E	$\underline{P}$					

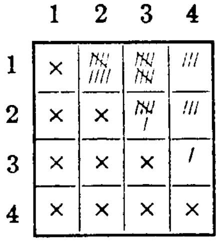


FIGURE 11c

Combination	Total number of comparisons	Number of coincidences		Discrepancy
		Expected	Observed	
				Percent
Messages 1 and 4	96	About 6	3	−50
Messages 2 and 4	96	About 6	3	−50
Messages 3 and 4	96	About 6	2	−83
Messages 1, 2, 3 and 4	582	About 39	33	−18

(6) It is believed that the procedure has been explained with sufficient detail to make further examples unnecessary. The student should bear in mind always that as he adds messages to the superimposition diagram it is necessary that he recalculate the number of comparisons so that the correct expected or theoretical number of coincidences will be before him to compare with the observed number. In adding messages he should see that the results of the separate tests are consistent, as well as those for the combined tests, otherwise he may be led astray at times by the overbalancing effect of the large number of coincidences for the already ascertained, correct superimpositions.

44. Subsequent steps.—*a.* In paragraph 43*a* four messages were given of a series supposedly enciphered by a long keying sequence, and the succeeding paragraphs were devoted to an explanation of the preparatory steps in the solution. The messages have now been properly superimposed, so that the text has been reduced to monoalphabetic columnar form, and the matter is now to be pursued to its ultimate stages.

b. The four messages employed in the demonstration of the principles of the κ test have served their purpose. The information that they are messages enciphered by an intelligible running key, by reversed standard cipher alphabets, was withheld from the student, for pedagogical reasons. Were the key a random sequence of letters instead of intelligible text, the explanation of the coincidence test would have been unchanged in the slightest particular, so far as concerns the mechanics of the text itself. Were the cipher alphabets unknown, mixed alphabets, the explanation of the κ test would also have been unchanged in the slightest particular. But, as stated before, the four messages actually represent encipherments by means of an intelligible running key, by reversed standard alphabets; they will now be used to illustrate the solution of cases of this sort.

c. Assuming now that the cryptanalyst is fully aware that the enemy is using the running-key system with reversed standard alphabets (obsolete U. S. Army cipher disk), the method of solution outlined in paragraph 38 will be illustrated, employing the first of the four messages referred to above, that beginning PGLPN HUFRK SAUQQ. The word DIVISION will be taken as a probable word and tested against the key, beginning with the very first letter of the message. Thus:

Cipher text	P	G	L	P	N	H	U	F	R	K	S	A	U	Q	Q	.	.
Assumed plain text	D	I	V	I	S	I	O	N	.	.	.	.	.	.	.	.	.
Resultant key text	S	O	G	X	F	.	.	.	.	.	.	.	.	.	.	.	.

The resultant key text is unintelligible and the word DIVISION is shifted one letter to the right.

Cipher text	P	G	L	P	N	H	U	F	R	K	S	A	U	Q	Q	.	.
Assumed plain text	.	D	I	V	I	S	I	O	N	.	.	.	.	.	.	.	.
Resultant key text	.	J	T	K	.	.	.	.	.	.	.	.	.	.	.	.	.

Again the resultant key text is unintelligible and the hypothetical word DIVISION is shifted once more. Continuation of this process to the end of the message proves that the word is not present. Another probable word is assumed: REGIMENT. When the point shown below is reached, note the results:

Cipher text	P	G	L	P	N	H	U	F	R	K	S	A	U	Q	Q	.	.
Assumed plain text	.	.	.	.	R	E	G	I	M	E	N	T	.	.	.	.	.
Resultant key text	.	.	.	.	E	L	A	N	D	O	F	T	.	.	.	.	.

It certainly looks as though intelligible text were being obtained as key text. The words LAND OF T . . . suggest that THE be tried. The key letters HE give NO, making the plain text read . . . . REGIMENT NO. . . . The four spaces preceding REGIMENT suggest such words as HAVE, SEND, MOVE, THIS, etc. A clue may be found by assuming that the E before LAND in the key is part of the word THE. Testing it on the cipher text gives IS for the plain text, which certainly indicates that the message begins with the word THIS. The latter yields IN for the first two key letters. And so on, the process of checking one text against the other continuing until the entire message and the key text have been reconstructed.

d. Thus far the demonstration has employed but one of the four messages available for solution. When the reconstruction process is applied to all four simultaneously it naturally goes much faster, with reduced necessity for assuming words after an initial entering wedge has been driven into one message. For example, note what happens in this case just as soon as the word REGIMENT is tried in the proper place:

	Key text	.	.	.	.	E	L	A	N	D	O	F	T	.	.	.	. . .
No. 1	Cipher text	P	G	L	P	N	H	U	F	R	K	S	A	U	Q	Q	. . .
	Plain text	.	.	.	.	R	E	G	I	M	E	N	T	.	.	.	. . .
No. 2	Cipher text	.	.	.	C	W	H	P	K	K	X	F	L	U	M	K	. . .
	Plain text	.	.	.	.	I	E	L	D	T	R	A	I	.	.	.	. . .
No. 3	Cipher text		W	F	W	T	D	N	H	T	G	M	R	A	A	Z	. . .
	Plain text		.	.	.	L	I	N	G	K	I	T	C	.	.	.	. . .
No. 4	Cipher text			T	U	L	D	H	N	Q	E	Z	Z	U	T	Y	. . .
	Plain text			.	.	T	I	T	A	N	K	G	U	.	.	.	. . .

It is obvious that No. 2 begins with FIELD TRAIN; No. 3, with ROLLING KITCHEN; No. 4 with ANTITANK GUN. These words yield additional key letters, the latter suggest additional plain text, and thus the process goes on until the solution is completed.

e. But now suppose that the key text that has been actually employed in encipherment is not intelligible text. The process is still somewhat the same, only in this case one must have at least two messages in the same key. For instead of checking a hypothetical word (assumed to be present in one message) against the key, *the same kind of a check is made against the other message or messages.* Assume, for instance, that in the case just described the key text, instead of being intelligible text, were a series of letters produced by applying a rather complex transposition to an originally intelligible key text. Then if the word REGIMENT were assumed to be present in the proper place in message No. 1 the resultant key letters would yield an unintelligible sequence. But these key letters when applied to message No. 2 would nevertheless yield IELDTRAI; when applied to message No. 3, LINGKITC, and so on. In short, the *text of one message is checked against the text of another message or messages;* if the originally assumed word is correct, then plain text will be found in the other messages.[8]

---

[8] Perhaps this is as good a place as any to make some observations which are of general interest in connection with the running-key principle, and which have no doubt been the subject of speculation on the part of some students. Suppose a basic, unintelligible, random sequence of keying characters which is not derived from the interaction of two or more shorter keys and which *never repeats* is employed *but once* as a key for encipherment. Can a cryptogram enciphered in such a system be solved? The answer to this question must unqualifiedly be this: even if the cipher alphabets are known sequences, cryptanalytic science is certainly powerless to attack such a cryptogram. Furthermore, so far as can now be discerned, no method of attack is likely ever to be devised. Short of methods based upon the alleged phenomena of telepathy—the very objective existence of which is denied by most "sane" investigators today—it is impossible for the present author to conceive of any way of attacking such a cryptogram.

This is a case (and perhaps the only case) in which the impossibility of cryptanalysis is mathematically demonstrable. Two things are involved in a complete solution in mathematics: not only must a satisfactory (logical) answer to the problem be offered, but also it must be demonstrated that the answer offered is *unique*, that is, the only possible one. (The mistake is often made that the latter phase of what constitutes a valid solution is overlooked—and this is the basic error which numerous alleged Bacon-Shakespeare "cryptographers" commit.) To attempt to solve a cryptogram enciphered in the manner indicated is analogous to an attempt to find a unique solution for a single equation containing two unknowns, with absolutely no data available for solution other than those given by that equation itself. It is obvious that no unique solution is possible in such a case, since *any one quantity whatsoever* may be chosen for one of the unknowns and the other will follow as a consequence. Therefore an infinite number of different answers, all equally valid, is possible. In the case of a *f.* All the foregoing work is, of course, based upon a knowledge of the cipher alphabets employed in the encipherment. What if the latter are unknown sequences? It may be stated at once that not much could be done with but four messages, even after they had been superimposed correctly, for the most that one would have in the way of data for the solution of the individual columns of text would be four letters per alphabet—which is not nearly enough. Data for solution by indirect symmetry by the detection of isomorphs cannot be expected, for no isomorphs are produced in this system. Solution can be reached only if there is sufficient text to permit of the analysis of the columns of the superimposition diagram. When there is this amount of text there are also repetitions which afford bases for the assumption of probable words. Only then, and after the values of a few cipher letters have been established can indirect symmetry be applied to facilitate the reconstruction of the primary components—if used.

g. Even when the volume of text is great enough so that each column contains say 15 to 20 letters, the problem is still not an easy one. But frequency distributions with 15 to 20 letters can usually be studied statistically, so that if two distributions present similar characteristics, the latter may be used as a basis for combining distributions which pertain to the same cipher alphabet. The next section will be devoted to a detailed treatment of the implications of the last statement.

---

cryptogram enciphered in the manner indicated, there is the equivalent of an equation with two unknowns; the key is one of the unknowns, the plain text is the other. One may conjure up an infinite number of different plain texts and offer any one of them as a "solution." One may even perform the perfectly meaningless labor of reconstructing the "key" for this selected "solution"; but since there is no way of proving from the cryptogram itself, or from the reconstructed key (which is unintelligible) whether the "solution" so selected is *the* actual plain text, all of the infinite number of "solutions" are equally valid. Now since it is inherent in the very idea of cryptography as a practical art that there must and can be only *one* actual solution (or plain text), and since none of this infinite number of different solutions can be proved to be *the one and only* correct solution, therefore, our common sense rejects them one and all, and it may be said that a cryptogram enciphered in the manner indicated is absolutely impossible to solve.

It is perhaps unnecessary to point out that the foregoing statement is no longer true when the running key constitutes intelligible text, or if it is used to encipher more than one message, or if it is the secondary resultant of the interaction of two or more short primary keys which go through cycles themselves. For in these cases there is additional information available for the delimitation of one of the pair of unknowns, and hence a unique solution becomes possible.

Now although the running-key system described in the first paragraph represents the ultimate goal of cryptographic security and is the ideal toward which cryptographic experts have striven for a long time, there is a wide abyss to be bridged between the recognition of a theoretically perfect system and its establishment as a practical means of secret intercommunication. For the mere mechanical details involved in the production, reproduction, and distribution of such keys present difficulties which are so formidable as to destroy the effectiveness of the method as a system of secret intercommunication suitable for groups of correspondents engaged in a voluminous exchange of messages.

SECTION XII

THE "CROSS-PRODUCT SUM" OR "χ TEST"[1]

Paragraph
Preliminary remarks 45
The nature of the "Cross-product sum" or "χ (Chi) test" in cryptanalysis 46
Derivation of the χ test 47
Applying the χ test in matching distributions 48

45. Preliminary remarks.—*a*. The real purpose of making the coincidence test in cases such as that studied in the preceding section is to permit the cryptanalyst to arrange his data so as to circumvent the obstacle which the enemy, by adopting a complicated polyalphabetic scheme of encipherment, places in the way of solution. The essence of the matter is that by dealing individually with the respective columns of the superimposition diagram the cryptanalyst has arranged the polyalphabetic text so that it can be handled as though it were monoalphabetic. Usually, the solution of the latter is a relatively easy matter, especially if there is sufficient text in the columns, or if the letters within certain columns can be combined into single frequency distributions, or if some cryptographic relationship can be established between the columns.

b. It is obvious that merely ascertaining the correct relative positions of the separate messages of a series of messages in a superimposition diagram is only a means to an end, and not an end in itself. The purpose is, as already stated, to reduce the complex, heterogeneous, polyalphabetic text to simple, homogeneous, monoalphabetic text. But the latter can be solved only when there are sufficient data for the purpose—and that depends often upon the type of cipher alphabets involved. The latter may be the secondary alphabets resulting from the sliding of the normal sequence against its reverse, or a mixed component against the normal, and so on. The student has enough information concerning the various cryptanalytic procedures which may be applied, depending upon the circumstances, in reconstructing different types of primary components and no more need be said on this score at this point.

c. The student should, however, realize one point which has thus far not been brought specifically to his attention. Although the superimposition diagram referred to in the preceding subparagraph may be composed of many columns, there is often only a relatively small number of *different* cipher alphabets involved. For example, in the case of two primary components of 26 letters each there is a maximum of 26 secondary cipher alphabets. Consequently, it follows that in such a case if a superimposition diagram is composed of say 100 columns, certain of those columns must represent similar secondary alphabets. There may, and probably will be, no regularity of recurrence of these repeated secondaries, for they are used in a manner directly governed by the letters composing the words of the key text or the elements composing the keying sequence.

d. But the latter statement offers an excellent clue. It is clear that the number of times a given secondary alphabet is employed in such a superimposition diagram depends upon the composition of the key text. Since in the case of a running-key system using a book as a key the key text constitutes intelligible text, it follows that *the various secondary alphabets will be employed with frequencies which are directly related to the respective frequencies of occurrence of letters in normal plain text.* Thus, the alphabet corresponding to key letter E should be the most frequently used; the alphabet corresponding to key letter T should be next in frequency, and so on. From this it follows that instead of being confronted with a problem involving as many different secondary cipher alphabets as there are columns in the superimposition diagram, the cryptanalyst will usually have not over 26 such alphabets to deal with; and allowing for the extremely improbable repetitive use of alphabets corresponding to key letters J, K, Q, X, and Z, it is likely that the cryptanalyst will have to handle only about 19 or 20 secondary alphabets.

---

[1] The χ test, presented in this section, as well as the Φ test, presented in Section XIV, were first described in an important paper, *Statistical Methods in Cryptanalysis*, 1935, by Solomon Kullback, Ph. D., Associate Cryptanalyst, Signal Intelligence Service. I take pleasure in acknowledging my indebtedness to Dr. Kullback's paper for the basic material used in my own exposition of these tests, as well as for his helpful criticisms thereof while in manuscript.

e. Moreover, since the E secondary alphabet will be used most frequently and so on, it is possible for the cryptanalyst to study the various distributions for the columns of the superimposition diagram with a view to assembling those distributions which belong to the same cipher alphabet, thus making the actual determination of values much easier in the combined distributions than would otherwise be the case.

f. However, if the keying sequence does not itself constitute intelligible text, even if it is a random sequence, the case is by no means hopeless of solution—provided there is sufficient text within columns so that the columnar frequency distributions may afford indications enabling the cryptanalyst to amalgamate a large number of small distributions into a smaller number of larger distributions.

g. In this process of assembling or combining individual frequency distributions which belong to the same cipher alphabet, recourse may be had to a procedure merely alluded to in connection with previous problems, and designated as that of "matching" distributions. The next few paragraphs will deal with this important subject.

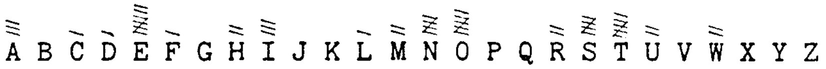
46. The nature of the "Cross-product sum" or "χ (Chi) test" in cryptanalysis.—*a:* The student has already been confronted with cases in which it was necessary or desirable to reduce a large number of frequency distributions to a smaller number by identifying and amalgamating distributions which belong to the same cipher alphabet. Thus, for example, in a case in which there are, say, 15 distributions but only, say, 5 separate cipher alphabets, the difficulty in solving a message can be reduced to a considerable degree provided that of the 15 distributions those which belong together can be identified and allocated to the respective cipher alphabets to which they apply.

b. This process of identifying distributions which belong to the same cipher alphabet involves a careful examination and comparison of the various members of the entire set of distributions to ascertain which of them present sufficiently similar characteristics to warrant their being combined into a single distribution applicable to one of the cipher alphabets involved in the problem. Now when the individual distributions are fairly large, say containing over 50 or 60 letters, the matter is relatively easy for the experienced cryptanalyst and can be made by the eye; but when the distributions are small, each containing a rather small number of letters, ocular comparison and identification of two or more distributions as belonging to the same alphabet become quite difficult and often inconclusive. In any event, the time required for the successful reduction of a multiplicity of individual small distributions to a few larger distributions is, in such cases, a very material factor in determining whether the solution will be accomplished in time to be of actual value or merely of historical interest.

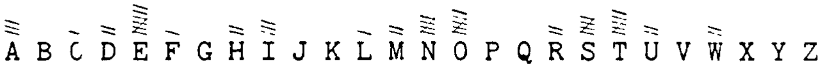
c. However, a certain statistical test, called the "cross-product sum" or "χ test", has been devised, which can be brought to bear upon this question and, by methods of mathematical comparison, eliminate to a large degree the uncertainties of the ocular method of matching and combining frequency distributions, thus in many cases materially reducing the time required for solution of a complex problem.

d. It is advisable to point out, however, that the student must not expect too much of a mathematical method of comparing distributions, because there are limits to the size of distributions to be matched below which these methods will not be effective. If two distributions contain some similar characteristics the mathematical method will merely afford a quantitative measure of the degree of similarity. Two distributions may actually pertain to the same cipher alphabet but, as occasionally happens, they may not present any external evidences of this relationship, in which case *no* mathematical method can indicate the fact that the two distributions are really similar and belong to the same alphabet.

47. Derivation of the χ test.—*a.* Consider the following plain-text distribution of 50 letters:



In a previous text[2] it was shown that the chance of drawing two identical letters in normal English telegraphic plain text is the sum of the squares of the relative probabilities of occurrence of the 26 letters in such text, which is .0667. That is, the probability of monographic coincidence in English telegraphic plain text is $\kappa_p = .0667$. In the message to which the foregoing distribution of 50 letters applies, the number of possible pairings (comparisons) that can be made between single letters is $\frac{50 \times 49}{2} = 1{,}225$. According to the theory of coincidences there should, therefore be $1{,}225 \times .0667 = 81.7065$ or approximately 82 coincidences of single letters. Examining the distribution it is found that there are 83 coincidences, as shown below:



3+0+0+1+21+0+0+1+3+0+0+0+1+10+15+0+0+1+10+15+1+0+1+0+0+0=83

The actual number of coincidences agrees very closely with the theoretical number, which is of course to be expected, since the text to which the distribution applied has been indicated as being normal plain text.

b. In the foregoing simple demonstration, let the number of comparisons that can be made in the distribution be indicated symbolically by $\frac{N(N-1)}{2}$, where N=the total number of letters in the distribution. Then the expected number of coincidences may be written as $\frac{.0067N(N-1)}{2}$, which may then be rewritten as

$$\text{(I)} \qquad \frac{.0067N^2 - .0667N}{2}.$$

c. Likewise, if f_A represents the number of occurrences of A in the foregoing distribution, then the number of coincidences for the letter A may be indicated symbolically by $\frac{f_A(f_A-1)}{2}$. And similarly, the number of coincidences for the letter B may be indicated by $\frac{f_B(f_B-1)}{2}$, and so on down to $\frac{f_Z(f_Z-1)}{2}$. The total number of actual coincidences found in the distribution is, of course, the sum of $\frac{f_A(f_A-1)}{2} + \frac{f_B(f_B-1)}{2} + \ldots \frac{f_Z(f_Z-1)}{2}$. If the symbol f_θ is used to indicate any of the letters A, B, . . . Z, and the symbol Σ is used to indicate that the sum of all the elements that follow this sign is to be found, then the sum of the actual coincidences noted in the distribution may be indicated thus: $\sum\frac{f_\theta(f_\theta-1)}{2}$, which may be rewritten as

---

[2] *Military Cryptanalysis, Part II,* Appendix 2.

(II) $$\sum\frac{f_\theta^2-f_\theta}{2}$$

d. Now although derived from different sources, the two expressions labeled (I) and (II) above are equal, or should be equal, in normal plain text. Therefore, one may write:

$$\sum\frac{f_\theta^2-f_\theta}{2}=\frac{.0667N^2-.0667N}{2}$$

Simplifying this equation:

(III) $$\Sigma f_\theta^2-\Sigma f_\theta=.0667N^2-.0667N$$

e. Now $\Sigma f_\theta=N$.

Therefore, expression (III) may be written as

(IV) $$\Sigma f_\theta^2-N=.0667N^2-.0667N,$$

which on reduction becomes:

(V) $$\Sigma f_\theta^2=.0667N^2+.9333N$$

This equation may be read as "the sum of the squares of the absolute frequencies of a distribution is equal to .0667 times the square of the total number of letters in the distribution, plus .9333 times the total number of letters in the distribution." The letter S_2 is often used to replace the symbol Σf_θ^2.

f. Suppose two monoalphabetic distributions are thought to pertain to the same cipher alphabet. Now if they actually do belong to the same alphabet, and if they are correctly[3] combined into a single distribution, the latter must still be monoalphabetic in character. That is, again representing the individual letter frequencies in one of these distributions by the general symbol f_{θ_1} the individual letter frequencies in the other distribution by f_{θ_2}, and the total frequency in the first distribution by N_1, that in the second distribution by N_2, then

(VI) $$\Sigma(f_{\theta_1}+f_{\theta_2})^2=.0667(N_1+N_2)^2+.9333(N_1+N_2)$$

Expanding the terms of this equation:

(VII) $$\Sigma f_{\theta_1}^2+2\Sigma f_{\theta_1}f_{\theta_2}+\Sigma f_{\theta_2}^2=.0667(N_1^2+2N_1N_2+N_2^2)+.9333N_1+.9333N_2$$

But from equation (V):

$$\Sigma f_{\theta_1}^2=.0667N_1^2+.9333N_1 \text{ and}$$

$$\Sigma f_{\theta_2}^2=.0667N_2^2+.9333N_2,$$

so that equation (VII) may be rewritten thus:

$$.0667N_1^2+.9333N_1+2\Sigma f_{\theta_1}f_{\theta_2}+.0667N_2^2+.9333N_2=$$
$$.0667(N_1^2+2N_1N_2+N_2^2)+.9333N_1+.9333N_2$$

---

[3] By "correctly" is meant that the two distributions are slid relative to each other to their proper superimposition.

Reducing to simplest terms by cancelling out similar expressions:

$$2\Sigma f_{\theta_1} f_{\theta_2} = .0667(2N_1N_2), \text{ or}$$

(VIII) $$\frac{\Sigma f_{\theta_1} f_{\theta_2}}{N_1N_2} = .0667$$

g. The last equation thus permits of establishing an expected value for the sum of the products of the corresponding frequencies of the two distributions being considered for amalgamation. The cross-product sum or χ *test for matching two distributions is based upon equation (VIII).*

48. Applying the χ test in matching distributions.—*a.* Suppose the following two distributions are to be matched:

f_1.............. $\bar{A}\ \overset{\overline{\equiv}}{B}\ C\ \overset{\equiv}{D}\ E\ \bar{F}\ G\ H\ \bar{I}\ J\ K\ \bar{L}\ M\ N\ \bar{O}\ P\ Q\ \overset{\equiv}{R}\ \overset{=}{S}\ \overset{=}{T}\ \bar{U}\ V\ \bar{W}\ \overset{\equiv}{X}\ Y\ \overset{=}{Z}$

f_2.............. $A\ \overset{=}{B}\ C\ D\ E\ \overset{\equiv}{F}\ G\ H\ \bar{I}\ J\ \bar{K}\ L\ M\ \bar{N}\ \bar{O}\ P\ Q\ \overset{\equiv}{R}\ \bar{S}\ \bar{T}\ U\ V\ W\ X\ \bar{Y}\ \overset{=}{Z}$

Let the frequencies be juxtaposed, for convenience in finding the sum of the cross products. Thus:

f_{θ_1}..................	1	4	0	3	0	1	0	0	1	0	0	1	0	0	1	0	0	3	2	2	1	0	1	3	0	2	.....$N_1=26$
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
f_{θ_2}..................	0	2	0	0	0	3	0	0	1	0	1	0	0	1	1	0	0	3	1	1	0	0	0	0	1	2	.....$N_2=17$
$f_{\theta_1} f_{\theta_2}$..............	0	8	0	0	0	3	0	0	1	0	0	0	0	0	1	0	0	9	2	2	0	0	0	0	0	4	

In this case $\Sigma f_{\theta_1} f_{\theta_2} = 8+3+1+1+9+2+2+4 = 30$

$$N_1N_2 = 26 \times 17 = 442$$

$$\frac{\Sigma f_{\theta_1} f_{\theta_2}}{N_1N_2} = \frac{30}{422} = .0711$$

b. The fact that the quotient (.0711) agrees very closely with the expected value (.0667) means that the two distributions very probably belong together or are properly matched. Note the qualifying phrase "very probably." It implies that there is no certainty about this business of matching distributions by mathematical methods. The mathematics serve only as measuring devices, so to speak, which can be employed to measure the degree of similarity that exists.

c. Instead of dividing $\Sigma f_{\theta_1} f_{\theta_2}$ by N_1N_2 and seeing how closely the quotient approximates the value .0667 or .0385, one may set up an expected value for $\Sigma f_{\theta_1} f_{\theta_2}$ and compare it with the observed value. Thus, in the foregoing example .0667 $(N_1N_2) = .0667 \times 422 = 28.15$; the observed value of $\Sigma f_{\theta_1} f_{\theta_2}$ is 30 and therefore the agreement between the expected and the observed values is quite close, indicating that the two distributions are probably properly matched.

d. There are other mathematical or statistical tests for matching, in addition to the χ test. Moreover, it is possible to go further with the χ test and find a measure of the reliance that may be placed upon the value obtained; but these points will be left for future discussion in subsequent texts.

e. One more point will, however, here be added in connection with the χ test. Suppose the very same two distributions in subparagraph *a* are again juxtaposed, but with f_{θ_2} shifted one interval to the left of the position shown in the subparagraph of reference. Thus:

f_{θ_1}	1	4	0	3	0	1	0	0	1	0	0	1	0	0	1	0	0	3	2	2	1	0	1	3	0	2	$N_1=26$
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
f_{θ_2}	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	$N_2=17$
	2	0	0	0	3	0	0	1	0	1	0	0	1	1	0	0	3	1	1	0	0	0	0	1	2	0	

Here $\Sigma f_{\theta_1} f_{\theta_2}=2+3+2+3=10$ and $\frac{\Sigma f_{\theta_1} f_{\theta_2}}{N_1 N_2}=\frac{10}{442}=.0226$

The observed ratio (.0226) is so much smaller than the expected (.0667) that it can be said that if the two distributions pertain to the same primary components they are not properly superimposed. *In other words, the* χ *test may also be applied in cases where two or more frequency distributions must be shifted relatively in order to find their correct superimposition.* The theory underlying this application of the χ test is, of course, the same as before: two monoalphabetic distributions when properly combined will yield a single distribution which should still be monoalphabetic in character. In applying the χ test in such cases it may be necessary to shift two 26-element distributions to various superimpositions, make the χ test for each superimposition, and take as correct that one which yields the best value for the test.

f. The nature of the problem will, of course, determine whether the frequency distributions which are to be matched should be compared (1) by direct superimposition, that is, setting the A to Z tallies of one distribution directly opposite the corresponding tallies of the other distribution, as in subparagraph *a*, or (2) by shifted superimposition, that is, keeping the A to Z tallies of the first distribution fixed and sliding the whole sequence of tallies of the second distribution to various superimpositions against the first.

SECTION XIII

APPLYING THE CROSS-PRODUCT SUM OR χ TEST

Paragraph

Study of a situation in which the χ test may be applied........ 49

Solution of a progressive-alphabet system by means of the χ test........ 50

Alternative method of solution........ 51

49. Study of a situation in which the χ test may be applied.—*a.* A simple demonstration of how the χ-test is applied in matching frequency distributions may now be set before the student. The problem involved is the solution of cryptograms enciphered according to the progressive-alphabet system (par. 36*b*), with secondary alphabets derived from the interaction of two identical mixed primary components. It will be assumed that the enemy has been using a system of this kind and that the primary components are changed daily.

b. Before attacking an actual problem of this type, suppose a few minutes be devoted to a general analysis of its elements. It is here assumed that the primary components are based upon the HYDRAULIC . . . Z sequence and that the cipher component is shifted toward the right one step at a time. Consider a cipher square such as that shown in Figure 12, which is applicable to the type of problem under study. It has been arranged in the form of a deciphering square. In this square, *the horizontal sequences are all identical but merely shifted relatively; the letters inside the square are plain-text letters.*

ALPHABET No.

CIPHER LETTER	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R
B	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C
C	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I
D	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y
E	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	X	D	R	A	U	L	I	C	B
F	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E
G	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F
H	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z
I	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L
J	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G
K	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J
L	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U
M	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K
N	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M
O	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N
P	P	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O
Q	Q	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P
R	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D
S	S	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q
T	T	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S
U	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H	Y	D	R	A
V	V	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T
W	W	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V
X	X	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W
Y	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z	H
Z	Z	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X

[Plain-text letters are within the square proper]

FIGURE 12.

c. If, for mere purposes of demonstration, instead of letters within the cells of the square there are placed tallies corresponding in number with the normal frequencies of the letters occupying the respective cells, the cipher square becomes as follows (showing only the 1st three rows of the square):

ALPHABET No.

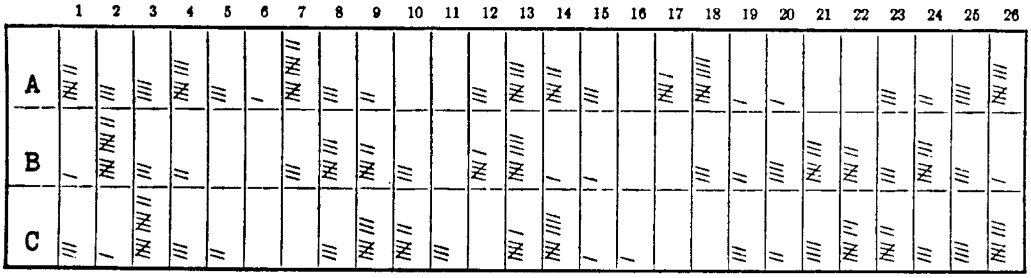

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
A
B
C

FIGURE 13a.

d. It is obvious that here is a case wherein if two distributions pertaining to the square are isolated from the square, the χ test (matching distributions) can be applied to ascertain how the distributions should be shifted relative to each other so that they can be superimposed and made to yield a monoalphabetic composite. There is obviously one correct superimposition out of 25 possibilities. In this case, the **B** row of tallies must be displaced 5 intervals to the right in order to match it and amalgamate it with the **A** row of tallies. Thus:

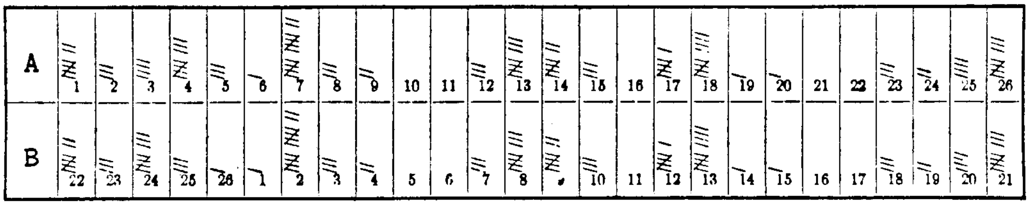

A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
B
22
23
24
25
26
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

FIGURE 13*b*.

e. Note that the amount of displacement, that is, the number of intervals the **B** sequence must be shifted to make it match the **A** sequence in Figure 13*b*, corresponds exactly to the distance between the letters A and B in the primary cipher component, which is 5 intervals. Thus: ... $\overset{0}{\mathrm{A}}\ \overset{1}{\mathrm{U}}\ \overset{2}{\mathrm{L}}\ \overset{3}{\mathrm{I}}\ \overset{4}{\mathrm{C}}\ \overset{5}{\mathrm{B}}$ ... . The fact that the primary plain component is in this case identical with the primary cipher component has nothing to do with the matter. *The displacement interval is being measured on the cipher component.* It is important that the student see this point very clearly. He can, if he like, prove the point by experimenting with two different primary components.

f. Assuming that a message in such a system is to be solved, the text is transcribed in rows of 26 letters. A uniliteral frequency distribution is made for each column of the transcribed text, the 26 separate distributions being compiled within a single square such as that shown in Figure 14. Such a square may be termed a *frequency distribution square.*

g. Now the vertical columns of tallies within such a distribution square constitute frequency distributions of the usual type: They show the distribution of the various cipher letters in each cipher alphabet. If there were many lines of text, all arranged in periods of 26 letters, then each column of the frequency square could be solved in the usual manner, by the application of the simple principles of monoalphabetic frequency. But what do the horizontal rows of tallies within the square represent? Is it not clear that the first such row, the **A** row, merely shows the distribution of A_c throughout the successive cipher alphabets? *And does not this graphic picture of the distribution of* A_c *correspond to the sequence of letters composing the primary plain component?* Furthermore, is it not clear that what has been said of the **A** row of tallies applies equally to the B, C, D, ... Z rows? Finally, is it not clear that the graphic pictures of all the distributions correspond to the *same* sequence of letters, except that the sequence begins with a different letter in each row? In other words, all the horizontal rows of tallies within the distribution square apply to the *same* sequence of plain-text letters, the sequences in one row merely beginning with a different letter from that with which another row begins. The sequences of letters to which the tallies apply in the various rows are merely displaced relative to one another. Now if there are sufficient data for statistical purposes in the various horizontal sequences of tallies within the distribution square, these sequences, being approximately similar, can be studied by means of the χ test *to find their relative displacements.* And in finding the latter a method is provided whereby the primary cipher component may be reconstructed, since the correct assembling of the displacement data will yield the sequence of letters constituting the primary cipher component. If the plain component is identical with the cipher component, the solution is immediately at hand; if they are different, the solution is but one step removed. Thus, there has been elaborated a method of solving this type of cipher system *without making any assumptions of values for cipher letters.*

50. Solution of a progressive-alphabet system by means of the χ test.—*a.* The following cryptogram has been enciphered according to the method indicated, by progressive, simple, uninterrupted shifting of a primary cipher component against an identical primary plain component.

CRYPTOGRAM

```
W G J J M   M M J X E   D G C O C   F T R P B   M I I I K   Z R Y N N
B U F R W   W W W Y O   I H F J K   O K H T T   A Z C L J   E P P F R
W C K O O   F F F G E   P Q R Y Y   I W X M X   U D I P F   E X M L L
W F K G Y   P B B X C   H B F Y I   E T X H F   B I V D I   P N X I V
R P W T M   G I M P T   E C J B O   K V B U Q   G V G F F   F K L Y Y
C K B I W   X M X U D   I P F F U   Y N V S S   I H R M H   Y Z H A U
Q W G K T   I U X Y J   J A O W Z   O C F T R   P P O Q U   S G Y C X
V C X U C   J L M L L   Y E K F F   Z V Q J Q   S I Y S P   D S B B J
U A H Y N   W L O C X   S D Q V C   Y V S I L   I W N J O   O M A Q S
L W Y J G   T V P Q K   P K T L H   S R O O N   I C F E V   M N V W N
B N E H A   M R C R O   V S T X E   N H P V B   T W K U Q   I O C A V
W B R Q N   F J V N R   V D O P U   Q R L K Q   N F F F Z   P H U R V
W L X G S   H Q W H P   J B C N N   J Q S O Q   O R C B M   R R A O N
R K W U H   Y Y C I W   D G S J C   T G P G R   M I Q M P   S G C T N
M F G J X   E D G C O   P T G P W   Q Q V Q I   W X T T T   C O J V A
A A B W M   X I H O W   H D E Q U   A I N F K   F W H P J   A H Z I T
W Z K F E   X S R U Y   Q I O V R   E R D J V   D K H I R   Q W E D G
E B Y B M   L A B J V   T G F F G   X Y I V G   R J Y E K   F B E P B
J O U A H   C U G Z L   X I A J K   W D V T Y   B F R U C   C C U Z Z
I N N D F   R J F M B   H Q L X H   M H Q Y Y   Y M W Q V   C L I P T
W T J Y Q   B Y R L I   T U O U S   R C D C V   W D G I G   G U B H J
V V P W A   B U J K N   F P F Y W   V Q Z Q F   L H T W J   P D R X Z
O W U S S   G A M H N   C W H S W   W L R Y Q   Q U S Z V   D N X A N
V N K H F   U C V V S   S S P L Q   U P C V V   V W D G S   J O G T C
H D E V Q   S I J P H   Q J A W F   R I Z D W   X X H C X   Y C T M G
U S E S N   D S B B K   R L V W R   V Z E E P   P P A T O   I A N E E
E E J N R   C Z B T B   L X P J J   K A P P M   J E G I K   R T G F F
H P V V V   Y K J E F   H Q S X J   Q D Y V Z   G R R H Z   Q L Y X K
X A Z O W   R R X Y K   Y G M G Z   B Y N V H   Q B R V F   E F Q L L
W Z E Y L   J E R O Q   S O Q K O   M W I O G   M B K F F   L X D X T
L W I L P   Q S E D Y   I O E M O   I B J M L   N N S Y K   X J Z J M
L C Z B M   S D J W Q   X T J V L   F I R N R   X H Y B D   B J U F I
R J I C T   U U U S K   K W D V M   F W T T J   K C K C G   C V S A G
Q B C J M   E B Y N V   S S J K S   D C B D Y   F P P V F   D W Z M T
B P V T T   C G B V T   Z K H Q D   D R M E Z   O O
```

b. The message is transcribed in lines of 26 letters, since that is the total number of secondary alphabets in the system. The transcribed text is shown below:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	W	G	J	J	M	M	M	J	X	E	D	G	C	O	C	F	T	R	P	B	M	I	I	I	K	Z
2	R	Y	N	N	B	U	F	R	W	W	W	W	Y	O	I	H	F	J	K	O	K	H	T	T	A	Z
3	C	L	J	E	P	P	F	R	W	C	K	O	O	F	F	F	G	E	P	Q	R	Y	Y	I	W	X
4	M	X	U	D	I	P	F	E	X	M	L	L	W	F	K	G	Y	P	B	B	X	C	H	B	F	Y
5	I	E	T	X	H	F	B	I	V	D	I	P	N	X	I	V	R	P	W	T	M	G	I	M	P	T
6	E	C	J	B	O	K	V	B	U	Q	G	V	G	F	F	F	K	L	Y	Y	C	K	B	I	W	X
7	M	X	U	D	I	P	F	F	U	Y	N	V	S	S	I	H	R	M	H	Y	Z	H	A	U	Q	W
8	G	K	T	I	U	X	Y	J	J	A	O	W	Z	O	C	F	T	R	P	P	O	Q	U	S	G	Y
9	C	X	V	C	X	U	C	J	L	M	L	L	Y	E	K	F	F	Z	V	Q	J	Q	S	I	Y	S
10	P	D	S	B	B	J	U	A	H	Y	N	W	L	O	C	X	S	D	Q	V	C	Y	V	S	I	L
11	I	W	N	J	O	O	M	A	Q	S	L	W	Y	J	G	T	V	P	Q	K	P	K	T	L	H	S
12	R	O	O	N	I	C	F	E	V	M	N	V	W	N	B	N	E	H	A	M	R	C	R	O	V	S
13	T	X	E	N	H	P	V	B	T	W	K	U	Q	I	O	C	A	V	W	B	R	Q	N	F	J	V
14	N	R	V	D	O	P	U	Q	R	L	K	Q	N	F	F	F	Z	P	H	U	R	V	W	L	X	G
15	S	H	Q	W	H	P	J	B	C	N	N	J	Q	S	O	Q	O	R	C	B	M	R	R	A	O	N
16	R	K	W	U	H	Y	Y	C	I	W	D	G	S	J	C	T	G	P	G	R	M	I	Q	M	P	S
17	G	C	T	N	M	F	G	J	X	E	D	G	C	O	P	T	G	P	W	Q	Q	V	Q	I	W	X
18	T	T	T	C	O	J	V	A	A	A	B	W	M	X	I	H	O	W	H	D	E	Q	U	A	I	N
19	F	K	F	W	H	P	J	A	H	Z	I	T	W	Z	K	F	E	X	S	R	U	Y	Q	I	O	V
20	R	E	R	D	J	V	D	K	H	I	R	Q	W	E	D	G	E	B	Y	B	M	L	A	B	J	V
21	T	G	F	F	G	X	Y	I	V	G	R	J	Y	E	K	F	B	E	P	B	J	O	U	A	H	C
22	U	G	Z	L	X	I	A	J	K	W	D	V	T	Y	B	F	R	U	C	C	C	U	Z	Z	I	N
23	N	D	F	R	J	F	M	B	H	Q	L	X	H	M	H	Q	Y	Y	Y	M	W	Q	V	C	L	I
24	P	T	W	T	J	Y	Q	B	Y	R	L	I	T	U	O	U	S	R	C	D	C	V	W	D	G	I
25	G	G	U	B	H	J	V	V	P	W	A	B	U	J	K	N	F	P	F	Y	W	V	Q	Z	Q	F
26	L	H	T	W	J	P	D	R	X	Z	O	W	U	S	S	G	A	M	H	N	C	W	H	S	W	W
27	L	R	Y	Q	Q	U	S	Z	V	D	N	X	A	N	V	N	K	H	F	U	C	V	V	S	S	S
28	P	L	Q	U	P	C	V	V	V	W	D	G	S	J	O	G	T	C	H	D	E	V	Q	S	I	J
29	P	H	Q	J	A	W	F	R	I	Z	D	W	X	X	H	C	X	Y	C	T	M	G	U	S	E	S
30	N	D	S	B	B	K	R	L	V	W	R	V	Z	E	E	P	P	P	A	T	O	I	A	N	E	E
31	E	E	J	N	R	C	Z	B	T	B	L	X	P	J	J	K	A	P	P	M	J	E	G	I	K	R
32	T	G	F	F	H	P	V	V	V	Y	K	J	E	F	H	Q	S	X	J	Q	D	Y	V	Z	G	R
33	R	H	Z	Q	L	Y	X	K	X	A	Z	O	W	R	R	X	Y	K	Y	G	M	G	Z	B	Y	N
34	V	H	Q	B	R	V	F	E	F	Q	L	L	W	Z	E	Y	L	J	E	R	O	Q	S	O	Q	K
35	O	M	W	I	O	G	M	B	K	F	F	L	X	D	X	T	L	W	I	L	P	Q	S	E	D	Y
36	I	O	E	M	O	I	B	J	M	L	N	N	S	Y	K	X	J	Z	J	M	L	C	Z	B	M	S
37	D	J	W	Q	X	T	J	V	L	F	I	R	N	R	X	H	Y	B	D	B	J	U	F	I	R	J
38	I	C	T	U	U	U	S	K	K	W	D	V	M	F	W	T	T	J	K	C	K	C	G	C	V	S
39	A	G	Q	B	C	J	M	E	B	Y	N	V	S	S	J	K	S	D	C	B	D	Y	F	P	P	V
40	F	D	W	Z	M	T	B	P	V	T	T	C	G	B	V	T	Z	K	H	Q	D	D	R	M	E	Z
41	O	O																								

c. A frequency distribution square is then compiled, each column of the text forming a separate distribution in columnar form in the square. The latter is shown in figure 14.

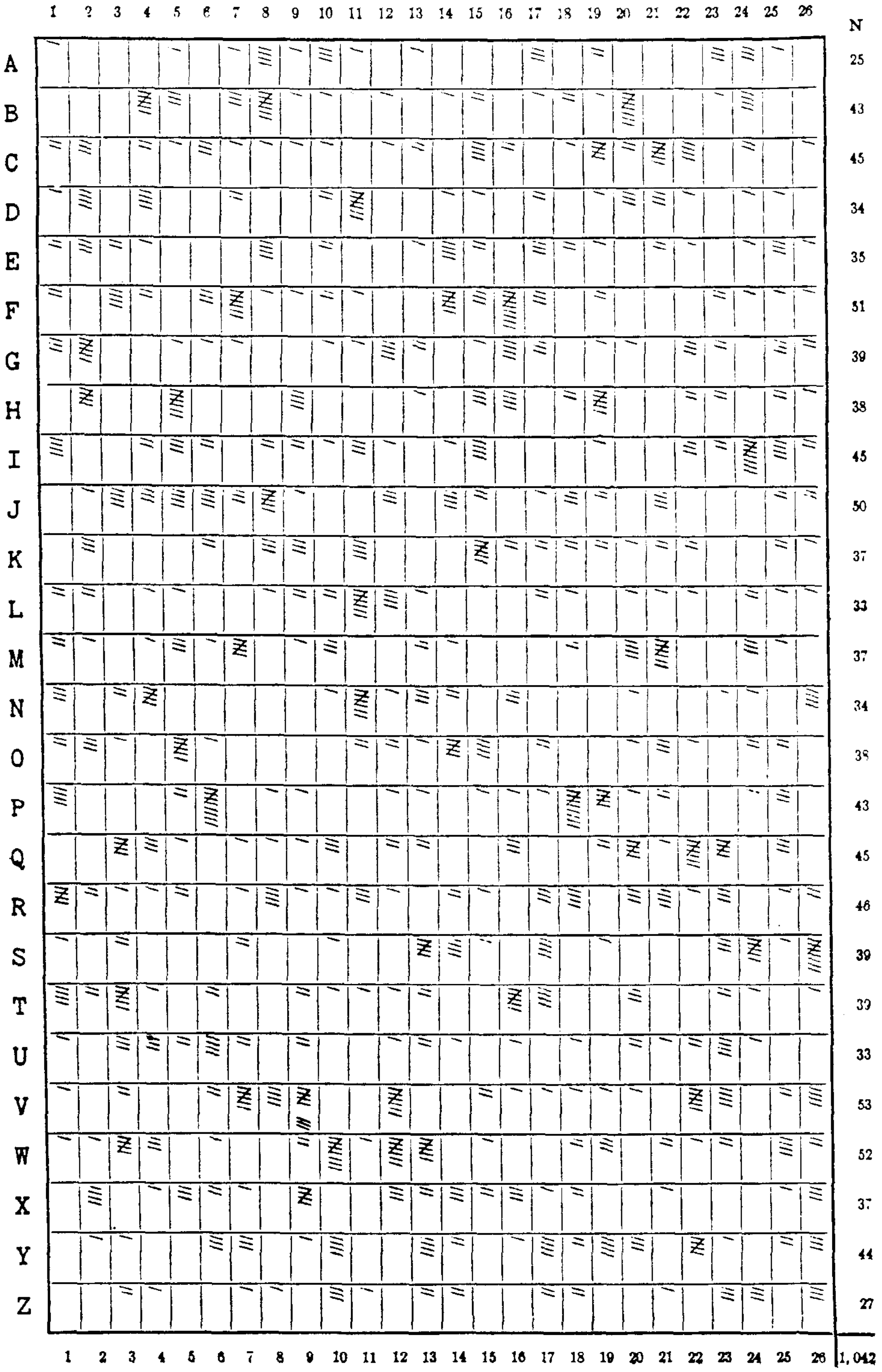

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
N
A 25
B 43
C 45
D 34
E 35
F 51
G 39
H 38
I 45
J 50
K 37
L 33
M 37
N 34
O 38
P 43
Q 45
R 46
S 39
T 39
U 33
V 53
W 52
X 37
Y 44
Z 27
1,042

FIGURE 14.

d. The χ test will now be applied to the horizontal rows of tallies in the distribution square, in accordance with the theory set forth in paragraph 49*g*. Since this test is purely statistical in character and becomes increasingly reliable as the size of the distributions increases, it is best to start by working with the two distributions having the greatest total numbers of tallies. These are the V and W distributions, with 53 and 52 occurrences, respectively. The results of three relative displacements of these two distributions are shown below, labeled "First test," "Second test," and "Third test."

FIRST TEST

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4	$N_V=53$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_W	24	25	26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
	0	4	2	1	1	5	3	0	1	0	0	2	8	1	7	6	0	1	0	0	2	3	0	2	1	2	$N_W=52$
f_Vf_W	0	0	4	0	0	10	18	0	8	0	0	14	0	0	14	6	0	1	0	0	0	18	0	0	2	8	$\Sigma f_Vf_W=103$

$$\frac{\Sigma f_Vf_W}{N_VN_W}=\frac{103}{2756}=.037$$

SECOND TEST

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4	$N_V=53$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_W	18	19	20	21	22	23	24	25	26	1	2	3	4	5	6	6	8	9	10	11	12	13	14	15	16	17	
	2	3	0	2	1	2	0	4	2	1	1	5	3	0	1	0	0	2	8	1	7	6	0	1	0	0	$N_W=52$
f_Vf_W	2	0	0	0	0	4	0	16	16	0	0	35	0	0	2	0	0	2	8	1	0	36	0	0	0	0	$\Sigma f_Vf_W=122$

$$\frac{\Sigma f_Vf_W}{N_VN_W}=\frac{122}{2756}=.044$$

THIRD TEST

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4	$N_V=53$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_W	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	
	3	0	1	0	0	2	8	1	7	6	0	1	0	0	2	3	0	2	1	2	0	4	2	1	1	5	$N_W=52$
f_Vf_W	3	0	2	0	0	4	48	4	56	0	0	7	0	0	4	3	0	2	1	2	0	24	8	0	2	20	$\Sigma f_Vf_W=190$

$$\frac{\Sigma f_Vf_W}{N_VN_W}=\frac{190}{2756}=.069$$

e. Since the last of the three foregoing tests gives a value somewhat better than the expected .0667, it looks as though the correct position of the W distribution with reference to the V distribution has been found. In practice, several more tests would be made to insure that other close approximations to .0667 will not be found, but these will here be omitted. The test indicates that the primary cipher component has the letters V and W in these positions: $\overset{1}{V}\ \overset{2}{.}\ \overset{3}{.}\ \overset{4}{W}$, since the correct superimposition requires that the 4th cell of the W distribution must be placed under the 1st cell of the V distribution (see the last superimposition above).

f. The next best distribution with which to proceed is the F distribution, with 51 occurrences. Paralleling the procedure outlined in paragraph 43, and for the same reasons, the F sequence is matched against the W and V sequences separately and then against both W and V sequences at their correct superimposition. The following shows the correct relative positions of the three distributions:

f_V	1	0	2	0	0	2	6	4	8	0	0	7	0	0	2	1	1	1	1	1	0	6	4	0	2	4	$N_V=53$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_F	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	4	5	6	7	
	1	1	2	1	0	0	6	3	9	3	0	2	0	0	0	2	1	1	1	2	0	4	2	0	3	7	$N_F=51$
f_Vf_F	1	0	4	0	0	0	36	12	72	0	0	14	0	0	0	2	1	1	1	2	0	24	8	0	6	28	$\Sigma f_Vf_F=212$

$$\frac{\Sigma f_Vf_F}{N_VN_F}=\frac{212}{2{,}703}=.078$$

---

f_W	1	1	5	3	0	1	0	0	2	8	1	7	6	0	1	0	0	2	3	0	2	1	2	0	4	2	$N_W=52$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_F	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	4	
	0	3	7	1	1	2	1	0	0	6	3	9	3	0	2	0	0	0	2	1	1	1	2	0	4	2	$N_F=51$
f_Wf_F	0	3	35	3	0	2	0	0	0	48	3	63	18	0	2	0	0	0	6	0	2	1	4	0	16	4	$\Sigma f_Wf_F=210$

$$\frac{\Sigma f_Wf_F}{N_WN_F}=\frac{210}{2{,}703}=.078$$

$f(V+W)$	4	0	3	0	0	4	14	5	15	6	0	8	0	0	4	4	1	3	2	3	0	10	6	1	3	9	$N_{V+W}=105$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
f_F	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	1	2	3	4	5	6	7	
	1	1	2	1	0	0	6	3	9	3	0	2	0	0	0	2	1	1	1	2	0	4	2	0	3	7	$N_F=51$
$f(V+W)f_F$	4	0	6	0	0	0	84	15	35	18	0	16	0	0	0	8	1	3	2	6	0	40	12	0	9	63	$\Sigma f_{(V+W)}f_F=422$

$$\frac{\Sigma f_{(V+W)}f_F}{N_{(V+W)}N_F}=\frac{422}{5{,}355}=.079$$

```
                             1 2 3 4 5 6 7 8
The test yields the sequence V . . W . . . F .
```

g. The process is continued in the foregoing manner until the entire primary cipher component has been reconstructed. It is obvious that as the work progresses the cryptanalyst is forced to employ smaller and smaller distributions, so that statistically the results are apt to become less and less certain. But to counterbalance this there is the fact that the number of possible superimpositions becomes progressively smaller as the work progresses. For example, at the commencement of operations the number of possible points for superimposing a second sequence against the first is 25; after the relative positions of 5 distributions have been ascertained and a 6th distribution is to be placed in the primary sequence being reconstructed, there are 21 possible positions; after the relative positions of 20 distributions have been ascertained, there are only 6 possible positions for the 21st distribution, and so on.

h. In the foregoing case the completely reconstructed primary cipher component is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K

Since it was stated that the problem involves identical primary components, both components are now at hand.

i. Of course, it is probable that in practical work the process of matching distributions would be interrupted soon after the positions of only a few letters in the primary component had been ascertained. For by trying partially reconstructed sequences on the cipher text the skeletons of some words would begin to show. By filling in these skeletons with the words suggested by them, the process of reconstructing the components is much facilitated and hastened.

j. The components having been reconstructed, only a moment or two is necessary to ascertain their initial position in enciphering the message. It is only necessary to juxtapose the two components so as to give "good" values for any one of the vertical distributions of Figure 14. This then gives the juxtaposition of the components for that column, and the rest follows very easily for the plain text may now be obtained by direct use of the components. The plain text of the message is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	W	G	J	J	M	M	M	J	X	E	D	G	C	O	C	F	T	R	P	B	M	I	I	I	K	Z
	W	I	T	H	T	H	E	I	M	P	R	O	V	E	M	E	N	T	S	I	N	T	H	E	A	I
2	R	Y	N	N	B	U	F	R	W	W	W	W	Y	O	I	H	F	J	K	O	K	H	T	T	A	Z
	R	P	L	A	N	E	A	N	D	T	H	E	M	E	A	N	S	O	F	C	O	M	M	U	N	I
3	C	L	J	E	P	P	F	R	W	C	K	O	O	F	F	F	G	E	P	Q	R	Y	Y	I	W	X
	C	A	T	I	O	N	A	N	D	W	I	T	H	T	H	E	V	A	S	T	S	I	Z	E	O	F
4	M	X	U	D	I	P	F	E	X	M	L	L	W	F	K	G	Y	P	B	B	X	C	H	B	F	Y
	M	O	D	E	R	N	A	R	M	I	E	S	S	T	R	A	T	E	G	I	C	S	U	R	P	R
5	I	E	T	X	H	F	B	I	V	D	I	P	N	X	I	V	R	P	W	T	M	G	I	M	P	T
	I	S	E	W	I	L	L	B	E	C	O	M	E	H	A	R	D	E	R	A	N	D	H	A	R	D
6	E	C	J	B	O	K	V	B	U	Q	G	V	G	F	F	F	K	L	Y	Y	C	K	B	I	W	X
	E	R	T	O	A	T	T	A	I	N	X	I	N	T	H	E	P	R	E	S	E	N	C	E	O	F
7	M	X	U	D	I	P	F	F	U	Y	N	V	S	S	I	H	R	M	H	Y	Z	H	A	U	Q	W
	M	O	D	E	R	N	A	V	I	A	T	I	O	N	A	N	D	F	A	S	T	M	O	V	I	N
8	G	K	T	I	U	X	Y	J	J	A	O	W	Z	O	C	F	T	R	P	P	O	Q	U	S	G	Y
	G	M	E	C	H	A	N	I	Z	E	D	E	L	E	M	E	N	T	S	G	R	E	A	T	E	R
9	C	X	V	C	X	U	C	J	L	M	L	L	Y	E	K	F	F	Z	V	Q	J	Q	S	I	Y	S
	C	O	M	P	L	E	X	I	T	I	E	S	M	O	R	E	S	U	B	T	L	E	D	E	C	E
10	P	D	S	B	B	J	U	A	H	Y	N	W	L	O	C	X	S	D	Q	V	C	Y	V	S	I	L
	P	T	I	O	N	S	S	T	R	A	T	E	G	E	M	S	A	N	D	F	E	I	N	T	S	W
11	I	W	N	J	O	O	M	A	Q	S	L	W	Y	J	G	T	V	P	Q	K	P	K	T	L	H	S
	I	L	L	H	A	V	E	T	O	B	E	E	M	P	L	O	Y	E	D	X	I	N	M	O	D	E
12	R	O	O	N	I	C	F	E	V	M	N	V	W	N	B	N	E	H	A	M	R	C	R	O	V	S
	R	N	W	A	R	F	A	R	E	I	T	I	S	S	T	I	L	L	P	O	S	S	I	B	L	E
13	T	X	E	N	H	P	V	B	T	W	K	U	Q	I	O	C	A	V	W	B	R	Q	N	F	J	V
	T	O	G	A	I	N	T	A	C	T	I	C	A	L	S	U	R	P	R	I	S	E	B	Y	M	A
14	N	R	V	D	O	P	U	Q	R	L	K	Q	N	F	F	F	Z	P	H	U	R	V	W	L	X	G
	N	Y	M	E	A	N	S	X	W	H	I	L	E	T	H	E	M	E	A	N	S	O	F	O	B	S
15	S	H	Q	W	H	P	J	B	C	N	N	J	Q	S	O	Q	O	R	C	B	M	R	R	A	O	N
	S	E	R	V	I	N	G	A	N	D	T	R	A	N	S	M	I	T	T	I	N	G	I	N	F	O

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
16	R	K	W	U	H	Y	Y	C	I	W	D	G	S	J	C	T	G	P	G	R	M	I	Q	M	P	S
	R	M	A	T	I	O	N	O	F	T	R	O	O	P	M	O	V	E	M	E	N	T	S	A	R	E
17	G	C	T	N	M	F	G	J	X	E	D	G	C	O	P	T	G	P	W	Q	Q	V	Q	I	W	X
	G	R	E	A	T	L	Y	I	M	P	R	O	V	E	D	O	V	E	R	T	H	O	S	E	O	F
18	T	T	T	C	O	J	V	A	A	A	B	W	M	X	I	H	O	W	H	D	E	Q	U	A	I	N
	T	H	E	P	A	S	T	T	H	E	M	E	C	H	A	N	I	C	A	L	M	E	A	N	S	O
19	F	K	F	W	H	P	J	A	H	Z	I	T	W	Z	K	F	E	X	S	R	U	Y	Q	I	O	V
	F	M	O	V	I	N	G	T	R	O	O	P	S	A	R	E	L	I	K	E	W	I	S	E	F	A
20	R	E	R	D	J	V	D	K	H	I	R	Q	W	E	D	G	E	B	Y	B	M	L	A	B	J	V
	R	S	P	E	E	D	I	E	R	X	A	L	S	O	F	A	L	S	E	I	N	F	O	R	M	A
21	T	G	F	F	G	X	Y	I	V	G	R	J	Y	E	K	F	B	E	P	B	J	O	U	A	H	C
	T	I	O	N	C	A	N	B	E	F	A	R	M	O	R	E	E	A	S	I	L	Y	A	N	D	Q
22	U	G	Z	L	X	I	A	J	K	W	D	V	T	Y	B	F	R	U	C	C	C	U	Z	Z	I	N
	U	I	C	K	L	Y	D	I	S	T	R	I	B	U	T	E	D	X	T	H	E	L	E	S	S	O
23	N	D	F	R	J	F	M	B	H	Q	L	X	H	M	H	Q	Y	Y	Y	M	W	Q	V	C	L	I
	N	T	O	B	E	L	E	A	R	N	E	D	F	R	O	M	T	H	E	O	P	E	N	I	N	G
24	P	T	W	T	J	Y	Q	B	Y	R	L	I	T	U	O	U	S	R	C	D	C	V	W	D	G	I
	P	H	A	S	E	O	F	A	L	L	E	N	B	Y	S	B	A	T	T	L	E	O	F	M	E	G
25	G	G	U	B	H	J	V	V	P	W	A	B	U	J	K	N	F	P	F	Y	W	V	Q	Z	Q	F
	G	I	D	O	I	S	T	H	A	T	S	U	R	P	R	I	S	E	I	S	P	O	S	S	I	B
26	L	H	T	W	J	P	D	R	X	Z	O	W	U	S	S	G	A	M	H	N	C	W	H	S	W	W
	L	E	E	V	E	N	I	N	M	O	D	E	R	N	W	A	R	F	A	R	E	B	U	T	O	N
27	L	R	Y	Q	Q	U	S	Z	V	D	N	X	A	N	V	N	K	H	F	U	C	V	V	S	S	S
	L	Y	B	Y	P	E	R	F	E	C	T	D	I	S	C	I	P	L	I	N	E	O	N	T	H	E
28	P	L	Q	U	P	C	V	V	V	W	D	G	S	J	O	G	T	C	H	D	E	V	Q	S	I	J
	P	A	R	T	O	F	T	H	E	T	R	O	O	P	S	A	N	D	A	L	M	O	S	T	S	U
29	P	H	Q	J	A	W	F	R	I	Z	D	W	X	X	H	C	X	Y	C	T	M	G	U	S	E	S
	P	E	R	H	U	M	A	N	F	O	R	E	T	H	O	U	G	H	T	A	N	D	A	T	T	E
30	N	D	S	B	B	K	R	L	V	W	R	V	Z	E	E	P	P	P	A	T	O	I	A	N	E	E
	N	T	I	O	N	T	O	D	E	T	A	I	L	O	N	T	H	E	P	A	R	T	O	F	T	H
31	E	E	J	N	R	C	Z	B	T	B	L	X	P	J	J	K	A	P	P	M	J	E	G	I	K	R
	E	S	T	A	F	F	B	A	C	K	E	D	U	P	B	Y	R	E	S	O	L	U	T	E	A	C
23	T	G	F	F	H	P	V	V	V	Y	K	J	E	F	H	Q	S	X	J	Q	D	Y	V	Z	G	R
	T	I	O	N	I	N	T	H	E	A	I	R	X	T	O	M	A	I	N	T	A	I	N	S	E	C
33	R	H	Z	Q	L	Y	X	K	X	A	Z	O	W	R	R	X	Y	K	Y	G	M	G	Z	B	Y	N
	R	E	C	Y	M	O	V	E	M	E	N	T	S	M	U	S	T	B	E	U	N	D	E	R	C	O
34	V	H	Q	B	R	V	F	E	F	Q	L	L	W	Z	E	Y	L	J	E	R	O	Q	S	O	Q	K
	V	E	R	O	F	D	A	R	K	N	E	S	S	A	N	D	C	O	V	E	R	E	D	B	I	V
35	O	M	W	I	O	G	M	B	K	F	F	L	X	D	X	T	L	W	I	L	P	Q	S	E	D	Y
	O	U	A	C	A	R	E	A	S	M	U	S	T	B	E	O	C	C	U	P	I	E	D	D	U	R

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
36	I	O	E	M	O	I	B	J	M	L	N	N	S	Y	K	X	J	Z	J	M	L	C	Z	B	M	S
	I	N	G	D	A	Y	L	I	G	H	T	H	O	U	R	S	X	U	N	O	B	S	E	R	V	E
37	D	J	W	Q	X	T	J	V	L	F	I	R	N	R	X	H	Y	B	D	B	J	U	F	I	R	J
	D	D	A	Y	L	I	G	H	T	M	O	V	E	M	E	N	T	S	W	I	L	L	R	E	Q	U
38	I	C	T	U	U	U	S	K	K	W	D	V	M	F	W	T	T	J	K	C	K	C	G	C	V	S
	I	R	E	T	H	E	R	E	S	T	R	I	C	T	I	O	N	O	F	H	O	S	T	I	L	E
39	A	G	Q	B	C	J	M	E	B	Y	N	V	S	S	J	K	S	D	C	B	D	Y	F	P	P	V
	A	I	R	O	B	S	E	R	V	A	T	I	O	N	B	Y	A	N	T	I	A	I	R	C	R	A
40	F	D	W	Z	M	T	B	P	V	T	T	C	G	B	V	T	Z	K	H	Q	D	D	R	M	E	Z
	F	T	A	R	T	I	L	L	E	R	Y	A	N	D	C	O	M	B	A	T	A	V	I	A	T	I
41	O	O																								
	O	N																								

WITH THE IMPROVEMENTS IN THE AIRPLANE AND THE MEANS OF COMMUNICATION AND WITH THE VAST SIZE OF MODERN ARMIES STRATEGIC SURPRISE WILL BECOME HARDER AND HARDER TO ATTAIN X IN THE PRESENCE OF MODERN AVIATION AND FAST MOVING MECHANIZED ELEMENTS GREATER COMPLEXITIES MORE SUBTLE DECEPTIONS STRATEGEMS AND FEINTS WILL HAVE TO BE EMPLOYED X IN MODERN WARFARE IT IS STILL POSSIBLE TO GAIN TACTICAL SURPRISE BY MANY MEANS X WHILE THE MEANS OF OBSERVING AND TRANSMITTING INFORMATION OF TROOP MOVEMENTS ARE GREATLY IMPROVED OVER THOSE OF THE PAST THE MECHANICAL MEANS OF MOVING TROOPS ARE LIKEWISE FAR SPEEDIER X ALSO FALSE INFORMATION CAN BE FAR MORE EASILY AND QUICKLY DISTRIBUTED X THE LESSON TO BE LEARNED FROM THE OPENING PHASE OF ALLENBYS BATTLE OF MEGGIDO IS THAT SURPRISE IS POSSIBLE EVEN IN MODERN WARFARE BUT ONLY BY PERFECT DISCIPLINE ON THE PART OF THE TROOPS AND ALMOST SUPERHUMAN FORETHOUGHT AND ATTENTION TO DETAIL ON THE PART OF THE STAFF BACKED UP BY RESOLUTE ACTION IN THE AIR X TO MAINTAIN SECRECY MOVEMENTS MUST BE UNDER COVER OF DARKNESS AND COVERED BIVOUAC AREAS MUST BE OCCUPIED DURING DAYLIGHT HOURS X UNOBSERVED DAYLIGHT MOVEMENTS WILL REQUIRE THE RESTRICTION OF HOSTILE AIR OBSERVATION BY ANTIAIRCRAFT ARTILLERY AND COMBAT AVIATION.

k. The student should clearly understand the real nature of the matching process employed to such good advantage in this problem. In practically all the previous cases frequency distributions were made of *cipher letters* occurring in a cryptogram, and the tallies in those distributions represented the actual occurrences of cipher letters. Furthermore, when these distributions were compared or matched, what were being compared were actually cipher alphabets. That is, the text was arranged in a certain way, so that letters belonging to the same cipher alphabet actually fell within the same column and the frequency distribution for a specific cipher alphabet was made by tabulating the letters in that column. Then if any distributions were to be compared, usually the entire distribution applicable to one cipher alphabet was compared with the entire distribution applying to another cipher alphabet. But in the problem just completed, what were compared in reality were not frequency distributions applying to the *columns* of the cipher text as transcribed on p. 83, but graphic representations of the variations in the frequencies of *plain-text letters falling in identical sequences, the identities of these plain-text letters being unknown for the moment.* Only after the reconstruction has been completed do their identities become known, when the plain text of the cryptogram is established.

51. Alternative method of solution.—*a.* The foregoing method of solution is, of course, almost entirely statistical in nature. There is, however, another method of attack which should be brought to notice because in some cases the statistical method, involving the study of relatively large distributions, may not be feasible for lack of sufficient text. Yet in these cases there may be sufficient data in the respective alphabets to permit of some assumptions of values of cipher letters, or there may be good grounds for applying the probable-word method. The present paragraph will therefore deal with a method of solving progressive cipher systems which is based upon the application of the principles of indirect symmetry to certain phenomena arising from the mechanics of the progressive encipherment method itself.

b. Take the two sequences below and encipher the phrase FIRST BATTALION by the progressive method, sliding the cipher component to the left one interval after each encipherment.

COMPONENTS

Plain	H	Y	D	R	A	U	L	I	C	B	E	F	G	J	K	M	N	O	P	Q	S	T	V	W	X	Z
Cipher	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X

MESSAGE

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Plain	F	I	R	S	T	B	A	T	T	A	L	I	O	N
Cipher	E	I	C	N	X	D	S	P	Y	T	U	K	Y	Y

c. Certain letters are repeated in both plain text and cipher text. Consider the former. There are two I's, three T's, and two A's. Their encipherments are isolated below, for convenience in study.

	F	I	R	S	T	B	A	T	T	A	L	I	O	N	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Plain	.	I	.	.	.	.	.	.	.	.	.	I	.	.	(1)
Cipher	.	I	.	.	.	.	.	.	.	.	.	K	.	.	(2)
Plain	.	.	.	.	T	.	.	T	T	.	.	.	.	.	(3)
Cipher	.	.	.	.	X	.	.	P	Y	.	.	.	.	.	(4)
Plain	.	.	.	.	.	.	A	.	.	A	.	.	.	.	(5)
Cipher	.	.	.	.	.	.	S	.	.	T	.	.	.	.	(6)

The two I's in line (1) are 10 letters apart; reference to the cipher component will show that the interval between the cipher equivalent of the first I_p (which happens to be I_c) and the second I_p (which is K_c) is 10. Consideration of the mechanics of the enciphering system soon shows why this is so: since the cipher component is displaced one step with each encipherment, two identical letters n intervals apart in the plain text must yield cipher equivalents which are n intervals apart in the cipher component. Examination of the data in lines (3) and (4), (5) and (6) will confirm this finding. Consequently, it would appear that in such a system the successful application of the probable-word method of attack, coupled within indirect symmetry, can quickly lead to the reconstruction of the cipher component.

d. Now consider the repeated cipher letters in the example under *b.* There happens to be only two cases of repetition, both involving Y's. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
.	.	.	.	.	.	.	.	T	.	.	.	O	N
.	.	.	.	.	.	.	.	Y	.	.	.	Y	Y

Reference to the plain component will show that the plain-text letters represented by the three Y's appear in the order N O . . . T, that is, reversed with respect to their order in the plain text. But the intervals between these letters is correct. Again a consideration of the mechanics of the enciphering system shows why this is so: since the cipher component is displaced one step with each encipherment, two identical letters *n* intervals apart in the cipher text must represent plain-text letters which are *n* intervals apart in the plain component. In the present case the direction in which these letters run in the plain component is opposite to that in which the cipher component is displaced. That is, if the cipher component is displaced toward the left, the values obtained from a study of repeated plain-text letters give letters which coincide in sequence (interval and direction) with the same letters in the cipher component; the values obtained from a study of repeated cipher-text letters give letters the order of which must be reversed in order to make these letters coincide in sequence (interval and direction) with the same letters in the plain component. If the cipher component is displaced toward the right, this relationship is merely reversed: the values obtained from a study of the repeated plain-text letters must be reversed in their order when placing them in the cipher component; those yielded by a study of the repeated cipher-text letters are inserted in the plain component in their original order.

e. Of course, if the primary components are identical sequences the data from the two sources referred to in subparagraphs *c* and *d* need not be kept separate but can be combined and made to yield the primary component very quickly.

f. With the foregoing principles as background, and given the following message, which is assumed to begin with COMMANDING GENERAL FIRST ARMY (probable-word method of attack), the data yielded by this assumed text are shown in Figure 15.

MESSAGE

IKMKI LIDOL WLPNM VWPXW DUFFT
FNIIG XGAMX CADUV AZVIS YNUNL etc., etc.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Assumed plain text	C	O	M	M	A	N	D	I	N	G	G	E	N	E	R	A	L	F	I	R	S	T	A	R	M	Y
Cipher	I	K	M	K	I	L	I	D	O	L	W	L	P	N	M	V	W	P	X	W	D	U	F	F	T	F

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1			I																							
2															K											
3													M													
4													K													
5	I																									
6														L												
7				I																						
8									D																	
9														O												
10							L																			
11							W																			
12					L																					
13														P												
14					N																					
15																		M								
16	V																									
17												W														
18						P																				
19									X																	
20																		W								
21																			D							
22																				U						
23	F																									
24																		F								
25													T													
26																									F	

Figure 15.

Analysis of the data afforded by Figure 15, in conjunction with the principles of indirect symmetry, yields the following partial components:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Plain	A	.	L	I	C	.	E	F	G	.	.	M	N	O	.	.	S	.	.	.	.	.	.	Y	D	R
Cipher	.	.	M	K	V	.	L	W	N	O	.	F	.	P	.	.	.	.	.	I	.	.	.	.	T	.
	D	.	.	.	.	.	.	.	.	.	.	X														

Setting the two partial components into juxtaposition so that $C_p = I_c$ (first encipherment) the 8th value, $I_p = D_c$, gives the position of D in the cipher component and permits the addition of X to it, these being two letters which until now could not be placed into position in the cipher component. With these two partial sequences it becomes possible now to decipher many other letters in the message, gaps being filled in from the context. For example, the first few letters after ARMY decipher as follows:

	1	2	3	4	5	6	7	8	9	10	11	12
Cipher...............	N	I	I	G	X	G	A	M	X	C	A	D
Plain.................	.	I	L	.	.	.	.	E	O	.	.	R

The word after ARMY is probably WILL. This leads to the insertion of the letter W in the plain component and G in the cipher component. In a short time both components can be completely established.

g. In passing, it may be well to note that in the illustrative message in paragraph 50*a* the very frequent occurrence of tripled letters (MMM, WWW, FFF, etc.) indicates the presence of a frequently used short word, a frequently used ending, or the like, the letters of which are sequent in the plain component. An astute cryptanalyst who has noted the frequency of occurrence of such triplets could assume the value THE for them, go through the entire text replacing all triplets by THE, and then, by applying the principles of indirect symmetry, build up the plain component in a short time. With that much as a start, solution of the entire message would be considerably simplified.

h. The principles elucidated in this paragraph may, of course, also be applied to cases of progressive systems in which the progression is by intervals greater than 1, and, with necessary modifications, to cases in which the progression is not regular but follows a specific pattern, such as 1–2–3, 1–2–3, . . . , or 2–5–7–3–1, 2–5–7–3–1, and so. The latter types of progression are encountered in certain mechanical cryptographs, the study of which will be reserved for future texts.

SECTION XIV

THE "MONOALPHABETICITY" OR "Φ TEST"

	Paragraph
Purpose of the Φ test	52
Derivation of the Φ test	53
Applying the Φ test	54

52. Purpose of the Φ (phi) test.—*a.* The student has noted that the χ test is based upon the general theory of coincidences and employs the probability constants κ_p and κ_r. There is one more test of a related nature which may be useful for him to understand and its explanation will be given in the succeeding paragraphs.

b. In paragraph 48*e* it was stated that two monoalphabetic distributions when correctly combined will yield a single distribution which should still be monoalphabetic in character. This question arises, therefore, in the student's mind: Is there a test whereby he can ascertain methematically whether a distribution is monoalphabetic or not, especially in the case of one which has relatively few data? Such a test has been devised and is termed the "Φ (phi) test."

53. Derivation of the Φ test.—*a.* Consider a monographic or uniliteral frequency distribution which is monoalphabetic in composition. If there is a total of N letters in the distribution, in a system in which there are n possible elements, then there is a possible total of $\frac{N(N-1)}{2}$ pairs of letters (for comparison purposes).

b. Let the symbol f_A represent the number of occurrences of A, f_B the number of occurrences of B, and so on to f_Z. With regard to the letter A then, there are $\frac{f_A(f_A-1)}{2}$ coincidences. (Again the combinations of f_A things taken two at a time.) With regard to the letter B, there are $\frac{f_B(f_B-1)}{2}$ coincidences, and so on up to $\frac{f_Z(f_Z-1)}{2}$ coincidences for the letter Z. Now it has been seen that according to the κ test, in $\frac{N(N-1)}{2}$ comparisons of letters forming the two members of pairs of letters in normal English plain text, there should be $\frac{\kappa_p N(N-1)}{2}$ coincidences, where κ_p is the probability of monographic coincidence for the language in question.

c. Now the expected value of $\frac{f_A(f_A-1)}{2}+\frac{f_B(f_B-1)}{2}+\dots+\frac{f_Z(f_Z-1)}{2}$ is equal to the theoretical number of coincidences to be expected in $\frac{N(N-1)}{2}$ comparisons of two letters, which for normal plain text is κ_p times $\frac{N(N-1)}{2}$ and for random text is κ_r times $\frac{N(N-1)}{2}$. That is, for plain text:

$$\text{Expected value of } \frac{f_A(f_A-1)}{2}+\frac{f_B(f_B-1)}{2}+\dots+\frac{f_Z(f_Z-1)}{2}=\kappa_p\times\frac{N(N-1)}{2}, \text{ or}$$

(IX) Expected value of $f_A(f_A-1)+f_B(f_B-1)+\dots+f_Z(f_Z-1)=\kappa_p N(N-1)$; and for random text:

$$\text{Expected value of } \frac{f_A(f_A-1)}{2}+\frac{f_B(f_B-1)}{2}+\dots+\frac{f_Z(f_Z-1)}{2}=\kappa_r\times\frac{N(N-1)}{2}, \text{ or}$$

(X) Expected value of $f_A(f_A-1)+f_B(f_B-1)+\ldots+f_Z(f_Z-1)=\kappa_r N(N-1)$.
If for the left-hand side of equations (IX) and (X) the symbol $E\ (\Phi)$ is used, then these equations become:

(XI) For plain text . . . $E(\Phi_p)=\kappa_p N(N-1)$

(XII) For random text . . . $E(\Phi_r)=\kappa_r N(N-1)$,

where $E(\Phi)$ means the average or expected value of the expression in the parenthesis, κ_p and κ_r are the probabilities of monographic coincidence in plain and in random text, respectively.

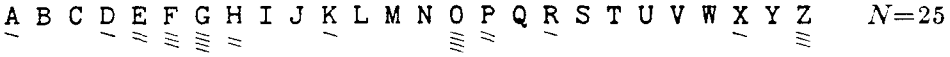
d. Now in normal English plain text it has been found that $\kappa_p=.0667$. For random text of a 26-letter alphabet $\kappa_r=.038$. Therefore, equations (XI) and (XII) may now be written thus:

(XIII) For normal English plain text . . . $E(\Phi_p)=.0667\ N(N-1)$

(XIV) For random text (26-letter alphabet) . . . $E(\Phi_r)=.0385\ N(N-1)$

e. By employing equations (XIII) and (XIV) it becomes possible, therefore, to test a piece of text for monoalphabeticity or for "randomness." That is, by using these equations one can mathematically test a very short cryptogram to ascertain whether it is a monoalphabetically enciphered substitution or involves several alphabets so that for all practical purposes it is equivalent to random text. This test has been termed the Φ test.

54. **Applying the Φ test.**—*a.* Given the following short piece of text, is it likely that it is normal English plain text enciphered monoalphabetically?



For this case the observed value of Φ is:

$$(1\times0)+(1\times0)+(2\times1)+(3\times2)+(4\times3)+(2\times1)+(1\times0)+(4\times3)+(2\times1)+(1\times0)+(1\times0)+(3\times2)=2+6+12+2+12+2+6=42$$

If this text were monoalphabetically enciphered English plain text the expected value of Φ is:

$$E(\Phi_p)=\kappa_p N(N-1)=.0667\times25\times24=40.0$$

If the text were random text, the expected value of Φ is:

$$E(\Phi_r)=\kappa_r N(N-1)=.0385\times25\times24=23.1$$

The conclusion is warranted, therefore, that the cryptogram is probably monoalphabetic substitution, since the observed value of Φ(42) more closely approximates the expected value for English plain text (40.0) than it does the expected value for random text (23.1). (As a matter of fact, the cryptogram was enciphered monoalphabetically.)

b. Here is another example. Given the following series of letters, does it represent a selection of English text enciphered monoalphabetically or does it more nearly represent a random selection of letters?

YOUIJ ZMMZZ MRNQC XIYTW RGKLH

The distribution and calculation are as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
			/				/	/	//	/	/	/	////	/	/		/	///		/	/		/	/	//	////
$f(f-1)$ . . .			0	0	0	2	0	0	0	6	0	0		0	2		0	0		0	0	2	6			

$$\Sigma f(f-1)=18 \text{ (That is, observed value of } \Phi=18)$$

$$E(\Phi_p)=.0667\times25\times24=40.0 \text{ (That is, expected value of } \Phi_p=40.0)$$

The conclusion is that the series of letters does not represent a selection of English text monoalphabetically enciphered. Whether or not it represents a random selection of letters cannot be told, but it may be said that if the letters actually do constitute a cryptogram, the latter is probably polyalphabetically enciphered. (As a matter of fact, the latter statement is true, for the message was enciphered by 25 alphabets used in sequence.)

c. The Φ test is, of course, closely related to the χ test and derives from the same general theory as the latter, which is that of coincidence. When two monoalphabetic distributions have been combined into a single distribution, the Φ test may be applied to the latter as a check upon the χ test. It is also useful in testing the columns of a superimposition diagram, to ascertain whether or not the columns are monoalphabetic.

SECTION XV

CONCLUDING REMARKS

Paragraph
Concluding remarks on aperiodic substitution systems........ 55
Synoptic table........ 56

55. Concluding remarks on aperiodic substitution systems.—*a*. The various systems described in the foregoing pages represent some of the more common and well-known methods of introducing complexities in the general scheme of cryptographic substitution with the view to avoiding or suppressing periodicity. There are, of course, other methods for accomplishing this purpose, which, while perhaps a bit more complex from a practical point of view, yield more desirable results from a cryptographic point of view. That is, these methods go deeper into the heart of the problem of cryptographic security and thus make the task of the enemy cryptanalyst much harder. But studies based on these more advanced methods will have to be postponed at this time, and reserved for a later text.

b. Thus far in these studies, aside from a few remarks of a very general nature, no attention has been paid to that other large and important class of ciphers, viz, transposition. It is desirable, before going further with substitution methods, that the student gain some understanding of how to solve certain of the more simple varieties of transposition ciphers. Consequently, in the text to succeed the present text, the student will temporarily lay aside the various useful methods and tools that he has been given for the solution of substitution ciphers and will turn his thoughts toward the methods of breaking down transposition ciphers.

56. Synoptic table.—Continuing the plan instituted in previous texts, of summarizing the textual material in the form of a very condensed chart called An Analytical Key for Military Cryptanalysis, the outline for the studies covered by Part III is shown on p. 119.

APPENDIX 1

ADDITIONAL NOTES ON METHODS FOR SOLVING PLAIN-TEXT AUTO-KEYED CIPHERS

Paragraph

Introductory remarks 1

Simple "mechanical" solution 2

Another "mechanical" solution 3

Solution of plain-text auto-keyed cryptograms when the introductory key is a word or phrase 4

Subsequent steps after determining the length of the introductory key 5

Conversion of foregoing aperiodic cipher into periodic form 6

Concluding remarks on auto-key systems 7

1. Introductory remarks.—*a.* In paragraph 33 of the text proper it was indicated that the method elucidated in paragraph 32 for solving plain-text auto-keyed ciphers is likely to be successful only if the cryptanalyst has been fortunate in his selection of a "probable word." Or, to put it another way, if the "probable words" which his imagination leads him to assume to be present in the text are really not present, then he is unfortunate, for solution will escape him. Hence, it is desirable to point out other principles and methods which are not so subject to chance. But because most of these methods are applicable only in special cases and because in general it is true that auto-key systems are no longer commonly encountered in practical military cryptography, it was thought best to exclude the exposition of these principles and methods from the text proper and to add them in an appendix, for the study of such students as find them of particular interest.

b. A complete discussion of the solution of plain-text auto-key systems, with examples, would require a volume in itself. Only one or two methods will be described, therefore, leaving the development of additional principles and methods to the ingenuity of the student who wishes to go more deeply into the subject. The discussion herein will be presented under separate headings, dependent upon the types of primary components employed.

c. As usual, the types of primary components may be classified as follows:

- (1) Primary components are identical.
    - (a) Both components progress in the same direction.
    - (b) Both components progress in opposite directions.
- (2) Primary components are different.

2. Simple "mechanical" solution.—*a.* (1) Taking up the case wherein the two identical primary components progress in the same direction, assume the following additional factors to be known by the cryptanalyst:

(a) The primary components are both normal sequences.

(b) The encipherment is by plain-text auto-keying.

(c) The enciphering equations are: $\Theta_{k/2}=\Theta_{i/1}$; $\Theta_{p/1}=\Theta_{c/2}$.

(2) A message beginning QVGLB TPJTF . . . is intercepted; the only unknown factor is the initial key letter. Of course, one could try to decipher the message using each key letter in turn, beginning with A and continuing until the correct key letter is tried, whereupon plain text will be obtained. But it seems logical to think that all the 26 possible "decipherments" might be derived from the first one, so that the process might be much simplified, and this is true, as

will now be shown. Taking the two cipher groups under consideration, let them be "deciphered" with initial key letter A:

Cipher.. QVGLBTPJTF
Deciphered with keyletter A.. QFBKRCNWXI

The deciphered text is certainly not "plain text." But if one completes the sequences initiated by these letters, using the direct standard sequence for the even columns, the reversed standard for the odd columns, the plain text sequence is seen to reappear on one generatrix: It is HOSTILE FOR(CE). From this it appears that instead of going through the labor of making 26 successive trials, which would consume considerable time, all that is necessary is to have a set of strips bearing the normal direct sequence and another set bearing the reversed normal sequence, and to align the strips, alternately direct and reversed, to the first "decipherment." The plain text will now reappear on one generatrix of the completion diagram. (See Fig. 1.)

Initial key letter	Q	V	G	L	B	T	P	J	T	F	
A	Q	F	B	K	R	C	N	W	X	I	
B	P	G	A	L	Q	D	M	X	W	J	
C	O	H	Z	M	P	E	L	Y	V	K	
D	N	I	Y	N	O	F	K	Z	U	L	
E	M	J	X	O	N	G	J	A	T	M	
F	L	K	W	P	M	H	I	B	S	N	
G	K	L	V	Q	L	I	H	C	R	O	
H	J	M	U	R	K	J	G	D	Q	P	
I	I	N	T	S	J	K	F	E	P	Q	
J	H	O	S	T	I	L	E	F	O	R	*
K	G	P	R	U	H	M	D	G	N	S	
L	F	Q	Q	V	G	N	C	H	M	T	
M	E	R	P	W	F	O	B	I	L	U	
N	D	S	O	X	E	P	A	J	K	V	
O	C	T	N	Y	D	Q	Z	K	J	W	
P	B	U	M	Z	C	R	Y	L	I	X	
Q	A	V	L	A	B	S	X	M	H	Y	
R	Z	W	K	B	A	T	W	N	G	Z	
S	Y	X	J	C	Z	U	V	O	F	A	
T	X	Y	I	D	Y	V	U	P	E	B	
U	W	Z	H	E	X	W	T	Q	D	C	
V	V	A	G	F	W	X	S	R	C	D	
W	U	B	F	G	V	Y	R	S	B	E	
X	T	C	E	A	U	Z	Q	T	A	F	
Y	S	D	D	I	T	A	P	U	Z	G	
Z	R	E	C	J	S	B	O	V	Y	H	

FIGURE 1.

b. The peculiar nature of the phenomenon just observed, *viz*, a completion diagram with the vertical sequences in adjacent columns progressing in opposite directions, those in alternate columns in the same direction, calls for an explanation. Although the matter seems rather mysterious, it will not be hard to understand. First, it is not hard to see why the letters in column 1 of Figure 1 should form the descending sequence QPO. . . for these letters are merely the ones resulting from the successive "decipherment" of Q_c by the successive key letters A, B, C, ... Now since the "decipherment" obtained from the 1st cipher letter in any row in Figure 1 becomes the key letter for "deciphering" the 2d cipher letter in the same row, it is apparent that as the letters in the 1st column progress in a reversed normal (descending) order, the letters in the 2d column *must* progress in a direct normal (ascending) order. The matter may perhaps become more clear if encipherment is regarded as a process of addition and decipherment as a process of subtraction. Instead of primary components or a Vigenère square, one may use simple arithmetic, assigning numerical values to the letters of the alphabet, beginning with A=0 and ending with Z=25. Thus on the basis of the pair of enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{c/2}$, the letter H_c enciphered by key letter M_k with direct primary components yields T_c. But using the following numerical values:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

the same result may be obtained thus: $H_p(M_k)=7+12=19=T_c$. Every time the number 25 is exceeded in the addition, one subtracts 26 from it and finds the letter equivalent for the remainder. In decipherment, the process is one of subtraction.[1] For example: $T_c(M_k=19-12=7=H_p$; $D_c(R_k)=3-17=[(26+3)-17]=29-17=12=M_p$. Using this arithmetical equivalent of normal sliding-strip encipherment, the phenomenon just noted can be set down in the form of a diagram (Fig. 2) which will perhaps make the matter clear.

---

[1] It will be noted that if the letters of the alphabet are numbered from 1 to 26, in the usual manner, the arithmetical method must be modified in a minor particular in order to obtain the same results as are given by employing the normal Vigenère square. This modification consists merely in subtracting 1 from the numerical value of the key letter. Thus:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

$$H_p(M_k)=8+(13-1)=8+12=20=T_c$$
$$T_c(M_k)=20-(13-1)=20-12=8=H_p$$

For an interesting extension of the basic idea involved in arithmetic cryptography, see:

Hill, Lester S. *Cryptography in an Algebraic Alphabet.* American Mathematical Monthly, Vol. XXXVI, No. 6, 1929.

Ibid. *Concerning certain linear transformation apparatus of cryptography.* American Mathematical Monthly, Vol. XXXVIII, No. 3, 1931.

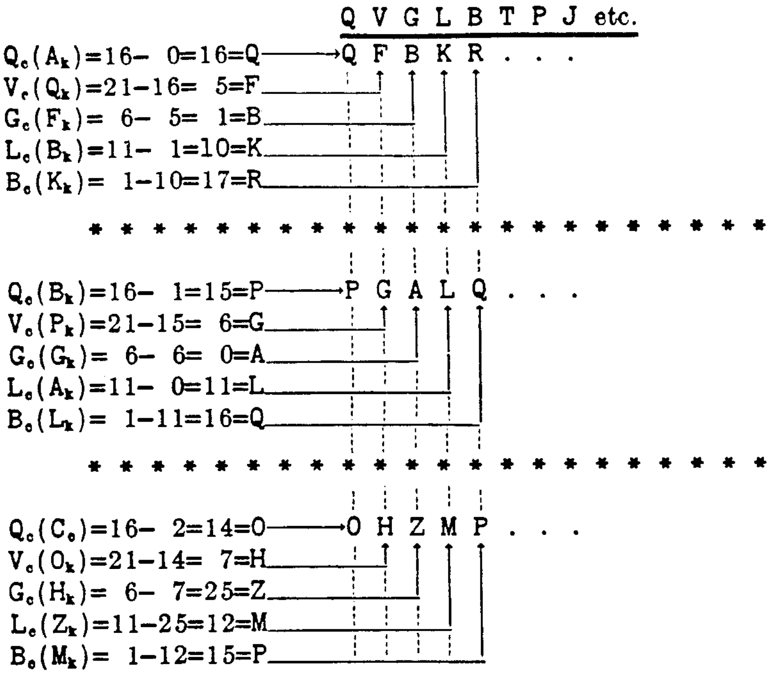

Q V G L B T P J etc.
$Q_c(A_k)=16-\ 0=16=Q$ → Q F B K R . . .
$V_c(Q_k)=21-16=\ 5=F$
$G_c(F_k)=\ 6-\ 5=\ 1=B$
$L_c(B_k)=11-\ 1=10=K$
$B_c(K_k)=\ 1-10=17=R$
* *
$Q_c(B_k)=16-\ 1=15=P$ → P G A L Q . . .
$V_c(P_k)=21-15=\ 6=G$
$G_c(G_k)=\ 6-\ 6=\ 0=A$
$L_c(A_k)=11-\ 0=11=L$
$B_c(L_k)=\ 1-11=16=Q$
* *
$Q_c(C_c)=16-\ 2=14=O$ → O H Z M P . . .
$V_c(O_k)=21-14=\ 7=H$
$G_c(H_k)=\ 6-\ 7=25=Z$
$L_c(Z_k)=11-25=12=M$
$B_c(M_k)=\ 1-12=15=P$

FIGURE 2.

Note how homologous letters of the three rows (joined by vertical dotted lines) form alternately descending and ascending normal sequences.

c. When the method of encipherment based upon enciphering equations $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/2}=\theta_{c/1}$ is used instead of the one based upon enciphering equations $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/1}=\theta_{c/2}$, the process indicated above is simplified by the fact that no alternation in the direction of the sequences in the completion diagram is required. For example:

Cipher	Y	H	E	B	P	D	T	B	J	D
Deciphered A=A	Y	F	J	K	Z	C	V	W	F	I
	Z	G	K	L	A	D	W	X	G	J
	A	H	L	M	B	E	X	Y	H	K
	B	I	M	N	C	F	Y	Z	I	L
	C	J	N	O	D	G	Z	A	J	M
	D	K	O	P	E	H	A	B	K	N
	E	L	P	Q	F	I	B	C	L	O
	F	M	Q	R	G	J	C	D	M	P
	G	N	R	S	H	K	D	E	N	Q
	*H	O	S	T	I	L	E	F	O	R

FIGURE 3.

d. (1) In the foregoing example the primary components were normal sequences, but the case of identical mixed components may be handled in a similar manner. Note the following example, based upon the following primary component (which is assumed to have been reconstructed from previous work):

F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Message............ U S I N L Y Q E O P ... etc.

(2) First, the message is "deciphered" with the initial key-letter A, and then a completion diagram is established, using sliding strips bearing the mixed primary component, alternate strips bearing the reversed sequence. Note Figure 4, in which the plain text, HOSTILE FOR(CE), reappears on a single generatrix. Note also that whereas in Figure 1 the odd columns contain the primary sequence in the reversed order, and the even columns contain the sequence in the direct order, in Figure 4 the situation is reversed: the odd columns contain the primary sequence in the direct order, and the even columns contain the sequence in the reversed order. This point is brought to notice to show that it is immaterial whether the direct order is used for odd columns or for even columns; the *alternation in direction* is all that is required in this type of solution.

e. (1) There is next to be considered the case in which the two primary components progress in opposite directions [par. 1*c* (1) (b)]. Here is a message, known to have been enciphered by reversed standard alphabets, plain-text auto-keying having been followed:

X T W Z L X H Z R X

(2) The procedure in this case is exactly the same as before, except that it is not necessary to have any alternation in direction of the completion sequences, which may be either that of the plain component or the cipher component. Note the solution in Figure 5. Let the student ascertain why the alternation in direction of the completion sequences is not necessary in this case.

(3) In the foregoing case the alphabets were reversed standard, produced by the sliding of the normal sequence against its reverse. But the underlying principle of solution is the same even if a mixed sequence were used instead of the normal; so long as the sequence is known, the procedure to be followed is exactly the same as demonstrated in subparagraphs (1) and (2) hereof. Note the following solution:

MESSAGE

V D D N C T S E P A . . .

Plain component....... F B P Y R C Q Z I G S E H T D J U M K V A L W N O X
Cipher component..... X O N W L A V K M U J D T H E S G I Z Q C R Y P B F

Note here that the primary mixed sequence is used for the completion sequence and that the plain text, HOSTILE FOR(CE), comes out on one generatrix. It is immaterial whether the direct or reversed mixed component is used for the completion sequence, so long as *all* the sequences in the diagram progress in the same direction. (See Fig. 6.)

f. (1) There remains now to be considered only the case in which the two components are different mixed sequences. Let the two primary components be as follows:

Plain.................. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher................ F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

and the message:

C F U Y L V X U D J

```
U S I N L Y Q E O P
-------------------
W D A Y K E L U I A
N T L P V S W J G V
O H W B A G N D S K
X E N F L I O T E M
F S O X W Z X H H U
B G X O N Q F E T J
P I E N O C B S D D
Y Z B W X R P G J T
R Q P L F Y Y I U H
C C Y A B P R Z M E
Q R R V P B C Q K S
Z Y C K Y F Q C V G
I P Q M R X Z R A I
G B Z U C O I Y L Z
S F I J Q N G P W Q
E X G D Z W S B N C
H O S T I L E F O R*
T N E H G A H X X Y
D W H E S V T O F P
J L T S E K D N B P
U A D G H M J W P F
M V J I T U U L Y X
K K U Z D J M A R O
V M M Q I D K V C N
A U K C U T V K Q W
L J V R M H A M Z L
↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑
```

FIGURE 4.

```
X T W Z L X H Z R X
C J N O D G Z A J M
-------------------
D K O P E H A B K N
E L P Q F I B C L O
F M Q R G J C D M P
G N R S H K D E N Q
H O S T I L E F O R*
I P T U J M F G P S
J Q U V K N G H Q T
K R V W L O H I R U
L S W X M P I J S V
M T X Y N Q J K T W
N U Y Z O R K L U X
O V Z A P S L M V Y
P W A B Q T M N W Z
Q X B C R U N O X A
R Y C D S V O P Y B
S Z D E T W P Q Z C
T A E F U X Q R A D
U B F G V Y R S B E
V C G H W Z S T C F
W D H I X A T U D G
X E I J Y B U V E H
Y F J K Z C V W F I
Z G K L A D W X G J
A H L M B E X Y H K
B I M N C F Y Z I L
```

FIGURE 5.

```
V D D N C T S E P A
Z V C I Y U Q L V X
-------------------
I A Q G R M Z W A F
G L Z S C K I N L B
S W I E Q V G O W P
E N G H Z A S X N Y
H O S T I L E F O R*
T X E D G W H B X C
D F H J S N T P F Q
J B T U E O D Y B Z
U P D M H X J R P I
M Y J K T F U C Y G
K R U V D B M Q R S
V C M A J P K Z C E
A Q K L U Y V I Q H
L Z V W M R A G Z T
W I A N K C L S I D
N G L O V Q W E G J
O S W X A Z N H S U
X E N F L I O T E M
F H O B W G X D H K
B T X P N S F J T V
P D F Y O E B U D A
Y J B R X H P M J L
R U P C F T Y K U W
C M Y Q B D R V M N
Q K R Z P J C A K O
Z V C I Y U Q L V X
```

FIGURE 6.

(2) First "decipher" the message with any arbitrarily selected initial key letter, say A, and complete the plain component sequence in the first column (Fig. 7*a*).

Cipher	C	F	U	Y	L	V	X	U	D	J
Plain	L	F	Q	X	W	X	A	W	S	F
	M									
	N									
	O									
	P									
	Q									
	R									
	S									
	T									
	U									
	V									
	W									
	X									
	Y									
	Z									
	A									
	B									
	C									
	D									
	E									
	F									
	G									
	H									
	I									
	J									
	K									

Figure 7*a*.

Cipher	C	F	U	Y	L	V	X	U	D	J
Plain	L	F	Q	X	W	X	A	W	S	E
	M	J								
	N	D								
	O	C								
	P	Y								
	Q	U								
	R	W								
	S	Q								
	T	N								
	U	K								
	V	H								
	W	E								
	X	B								
	Y	X								
	Z	T								
	A	G								
	B	Z								
	C	V								
	D	M								
	E	P								
	F	A								
	G	R								
	H	O								
	I	S								
	J	L								
	K	I								

Figure 7*b*.

Cipher	C	F	U	Y	L	V	X	U	D	J
Plain	L	F	Q	X	W	X	A	W	S	E
	M	J	B	C						
	N	D	C	Y						
	O	C	L	I						
	P	Y	N	G						
	Q	U	A	J						
	R	W	U	N						
	S	Q	K	L						
	T	N	T	Q						
	U	K	Y	A						
	V	H	E	S						
	W	E	F	D						
	X	B	P	B						
	Y	X	R	Z						
	Z	T	D	P						
	A	E	H	R						
	B	Z	J	O						
	C	V	X	E						
	D	M	Z	W						
	E	P	O	F						
	F	A	W	H						
	G	R	M	M						
	*H	O	S	T						
	I	S	G							
	J	L	V							
	K	I	I							

Figure 7*c*.

Now prepare a strip bearing the cipher component *reversed*, and set it below the plain component so that $F_p=L_c$, a setting given by the 1st two letters of the spurious "plain text" recovered. Thus:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B

(3) Now opposite each letter of the completion sequence in column 1, write its plain-component equivalent, as given by the juxtaposed sequences above. This gives what is shown in Figure 7*b*. Then reset the two sequences (reversed cipher component and the plain component) so that $Q_p=F_c$ (to correspond with the 2d and 3d letters of the spurious plain text); write down the plain-component equivalents of the letters in column 2, forming column 3. Continue this process, scanning the generatrices from time to time, resetting the two components and finding equivalents from column to column, until it becomes evident on what generatrix the plain text is reappearing. In Figure 7*c* it is seen that the plain text generatrix is the one beginning HOST, and from this point on the solution may be obtained directly, by using the two primary components.

(4) When the plain component is also a mixed sequence (and different from the cipher component), the procedure is identical with that outlined in subparagraphs (1)–(3) above. The fact that the plain component in the preceding case is the normal sequence is of no particular significance in the solution, for it acts as a mixed sequence would act under similar circumstances. To demonstrate, suppose the two following components were used in encipherment of the message below:

Plain.............. W B V I G X L H Y A J Z M N F O R P E Q D S C T K U
Cipher.............. F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Message.......... B B V Z U D Q X J D . . .

To solve the message, "decipher" the text with any arbitrarily selected initial key letter and proceed exactly as in subparagraphs (2) and (3) above. Thus:

Cipher.......................... B B V Z U D Q X J D
"Plain" (θ_k=X).......... V Y R I Y Z E F O R

Note the completion diagram in Figure 8 which shows the word HOST... very soon in the process. From this point on the solution may be obtained directly, by using the two primary components.

B	B	V	Z	U	D	Q	X	J	D
V	Y	R	I	Y	Z	E	F	O	R
I	Q	N	J						
G	E	Y	G						
X	V	W	Z						
L	L	K	O						
*H	O	S	T						
Y	K	B							
A	H	H							
J	M	V							
Z	D	X							
M	J	G							
N	G	J							
F	B	E							
O	I	Z							
R	T	L							
P	U	I							
E	R	O							
Q	S	A							
D	N	C							
S	P	P							
C	C	F							
T	F	Q							
K	A	U							
U	Z	M							
W	X	D							

FIGURE 8.

3. Another "mechanical" solution.—*a.* Another "mechanical" solution for the foregoing cases will now be described because it presents rather interesting cryptanalytic sidelights. Take the message

REFERENCE HIS PREFERENCE IN REFERENCE
BOOKS AND REFERENCE CHARTS . . .

and encipher it by plain-text auto-key, with normal direct primary components, initial key setting $A_p=G_c$. Then note the underscored repetitions:

R E F E R E N C E H I S P R E F E R E N C E I N R E F E
X V J J V V R P G L P A H G V J J V V R P G M V E V J J

R E N C E B O O K S A N D R E F E R E N C E C H A R T S
V V R P G F P C Y C S N Q U V J J V V R P G G J H Y K L

b. Now suppose the message has been intercepted and is to be solved. The only unknown factor will be assumed to be the initial key letter. Let the message be "deciphered" by means of any initial key letter,[2] say A, and then note the underscored repetitions in the spurious plain text.

Cipher.............. X V J J V V R P G L P A H G V J J V V R P G M V E V J J
"Plain text"...... X Y L Y X Y T W K B O M V L K Z K L K H I Y O H X Y L Y

Cipher.............. V V R P G F P C Y C S N Q U V J J V V R P G G J H Y K L
"Plain text"...... X Y T W K V U I Q M G H J L K Z K L K H I Y I B G S S T

The original four 8-letter repetitions now turn out to be two different sets of 9-letter repetitions. This calls for an explanation. Let the spurious plain text, with its real plain text be transcribed as though one were dealing with a periodic cipher involving two alphabets, as shown in Figure 9. It will here be seen that the letters in column 1 are monoalphabetic, and so are those in column 2. In other words, an auto-key cipher, which is commonly regarded as a polyalphabetic, aperiodic cipher, has been converted into a 2-alphabet, periodic cipher, the individual alphabets of which are now monoalphabetic in nature. The two repetitions of X Y L Y X Y T W K represent encipherments of the word REFERENCE, in alphabets 1–2–1–2–1–2–1–2–1; the two repetitions of L K Z K L K H I Y likewise represent encipherments of the same word but in alphabets 2–1–2–1–2–1–2–1–2.

c. Later on it will be seen how this method of converting an auto-key cipher into a periodic cipher may be applied to the case where an introductory key word is used as the initial keying element instead of a single letter, as in the present case.

1–2	1–2	1–2	1–2
R E	E F	R E	E F
X Y	K Z	X Y	K Z
F E	E R	N C	E R
L Y	K L	T W	K L
R E	E N	E B	E N
X Y	K H	K V	K H
N C	C E	O O	C E
T W	I Y	U I	I Y
E H	I N	K S	C H
K B	O H	Q M	I B
I S	R E	A N	A R
O M	X Y	G H	G S
P R	F E	D R	T S
V L	L Y	J L	S T

FIGURE 9.

---

[2] Except the actual key letter or a letter 13 intervals from it. See subparagraph (7) below.

d. The student has probably already noted that the phenomena observed in this subparagraph are the same as those observed in subparagraph 2*b*. In the latter subparagraph it was seen that the direction of the sequences in alternate columns had to be reversed in order to bring out the plain text on one generatrix. If this reversal is not done, then obviously the plain text would appear on *two* generatrices, which is equivalent to having the plain text reduced to two monoalphabets.

e. When reciprocal components are employed, the spurious plain text obtained by "decipherment" with a key setting other than the actual one will be monoalphabetic throughout. Note the following encipherment (with initial key setting $A_p=G_c$, using a reversed standard sequence sliding against the direct standard) and its "decipherment" by setting these two components $A_p=A_c$.

Plain text.............. R E F E R E N C E H I S P R E F E R E N C E . . .
Cipher................... P N Z B N N R L Y X Z Q D Y N Z B N N R L Y . . .
Spurious plain text.. L Y Z Y L Y H W Y B C M J L Y Z Y L Y H W Y . . .

Here the spurious plain text is wholly monoalphabetic.

f. The reason for the exception noted in footnote 2 on page 106 now becomes clear. For if the actual initial key letter (G) were used, of course the decipherment yields the correct plain text; if a letter 13 intervals removed from G is used as the key letter, the cipher alphabet selected for the first "decipherment" is the reciprocal of the real initial cipher alphabet and thereafter all alternate cipher alphabets are reciprocal. Hence the spurious text obtained from such a "decipherment" must be monoalphabetic.

g. In the foregoing case the primary components were identical normal sequences progressing in the same direction. If they were mixed sequences the phenomena observed above would still hold true, and so long as the sequences are known, the indicated method of solution may be applied.

h. When the two primary components are known but differently mixed sequences, this method of solution is too involved to be practical. It is more practicable to try successive initial key letters, noting the plain text each time and resetting the strips until the correct setting has been ascertained, as will be evidenced by obtaining intelligible plain text.

4. **Solution of plain-text auto-keyed cryptograms when the introductory key is a word or phrase.**—*a.* In the foregoing discussion of plain-text auto-keying, the introductory key was assumed to consist of a single letter, so that the subsequent key letters are displaced one letter to the right with respect to the text of the message itself. But sometimes a word or phrase may serve this function, in which case the subsequent key is displaced as many letters to the right of the initial plain-text letter of the message as there are letters in the initial key. This will not, as a rule, interfere in any way with the application of the principles of solution set forth in paragraph 28 to that part of the cryptogram subsequent to the introductory key, and a solution by the probable-word method and the study of repetitions can be reached. However, it may happen that trial of this method is not successful in certain cryptograms because of the paucity of repetitions, or because of failure to find a probable word in the text. When the cipher alphabets are known there is another point of attack which is useful and interesting. The method consists in finding the length of the introductory key and then solving by frequency principles. Just how this is accomplished will now be explained.

b. Suppose that the introductory key word is HORSECHESTNUT, that the plain-text message is as below, and that identical primary components progressing in the same direction are used to encipher the message, by enciphering equation $\theta_{k/2}=\theta_{1/1}$; $\theta_{p/1}=\theta_{c/2}$. Let the components be the normal sequence. The encipherment is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Key	H	O	R	S	E	C	H	E	S	T	N	U	T	M	Y	L	E	F	T	F	L	A	N	K	I	S
Plain	M	Y	L	E	F	T	F	L	A	N	K	I	S	R	E	C	E	I	V	I	N	G	H	E	A	V
Cipher	T	M	C	W	J	V	M	P	S	G	X	C	L	D	C	N	I	N	O	N	Y	G	U	O	I	N
Key	R	E	C	E	I	V	I	N	G	H	E	A	V	Y	A	R	T	I	L	L	E	R	Y	F	I	R
Plain	Y	A	R	T	I	L	L	E	R	Y	F	I	R	E	E	N	E	M	Y	I	S	M	A	S	S	I
Cipher	P	E	T	X	Q	G	T	R	X	F	J	I	M	C	E	E	X	U	J	T	W	D	Y	X	A	Z
Key	E	E	N	E	M	Y	I	S	M	A	S	S	I	N	G	T	R	O	O	P	S	T	O	L	E	F
Plain	N	G	T	R	O	O	P	S	T	O	L	E	F	T	F	R	O	N	T	A	N	D	C	O	N	C
Cipher	R	K	G	V	A	M	X	K	F	O	D	W	N	G	L	K	F	B	H	P	F	W	Q	Z	R	H
Key	T	F	R	O	N	T	A	N	D	C	O	N	C	E	N	T	R	A	T	I	N	G	A	R	T	I
Plain	E	N	T	R	A	T	I	N	G	A	R	T	I	L	L	E	R	Y	T	H	E	R	E	X	W	I
Cipher	X	S	K	F	N	M	I	A	J	C	F	G	K	P	Y	X	I	Y	M	P	R	X	E	O	P	Q
Key	L	L	E	R	Y	T	H	E	R	E	X	W	I	L	L	N	E	E	D	C	O	N	S	I	D	E
Plain	L	L	N	E	E	D	C	O	N	S	I	D	E	R	A	B	L	E	R	E	I	N	F	O	R	C
Cipher	W	W	R	V	C	W	J	S	E	W	F	Z	M	C	L	O	P	I	U	G	W	A	X	W	U	G
Key	R	A	B	L	E	R	E	I	N	F	O	R	C	E	M	E	N	T	S	T	O	M	A	I	N	T
Plain	E	M	E	N	T	S	T	O	M	A	I	N	T	A	I	N	M	Y	P	O	S	I	T	I	O	N
Cipher	V	M	F	Y	X	J	X	W	Z	F	W	E	V	E	U	R	Z	R	H	H	G	U	T	Q	B	G

It will now be noted that since the introductory key contains 13 letters the 14th letter of the message is enciphered by the 1st letter of the plain text, the 15th by the 2d, and so on. Likewise, the 27th letter is enciphered by the 14th, the 28th by the 15th, and so on. Hence, if the 1st cipher letter is deciphered, this will give the key for deciphering the 14th, the latter will give the key for the 27th, and so on. An important step in the solution of a message of this kind would therefore involve ascertaining the length of the introductory key. This step will now be explained.

c. Since the plain text itself constitutes the key letters in this system (after the introductory key), these key letters will occur with their normal frequencies, and this means that there will be many occurrences of E, T, O, A, N, I, R, S, enciphered by E_k; there will be many occurrences of these same high-frequency letters enciphered by T_k, by O_k, by A_k, and so on. In fact, the number of times each of these combinations will occur may be calculated statistically. With the enciphering conditions set forth under *b* above, E_p enciphered by T_k, for example, will yield the same cipher equivalent as T_p enciphered by E_k; in other words two encipherments of any pair of letters of which either may serve as the key for enciphering the other must yield the same cipher resultant.[3] It is the cryptographic effect of these two phenomena working together which permits of ascertaining the length of the introductory key in such a case. For every time a given letter, θ_p, occurs in the plain text it will occur *n* letters later as a key letter, θ_k, and *n* in this case equals the length of the introductory key. Note the following illustration:

---

[3] It is important to note that the two components must be identical sequences and progress in the same direction. If this is not the case, the entire reasoning is inapplicable.

	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
(1) Key	H	O	R	S	E	C	H	E	S	T	N	U	T	.	.	.	.	.	T	.	.	.	.	.	.	.
(2) Plain	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.
(3) Cipher	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.	.

	1	2	3	4	5	6	7	8	9
(1) Key	.	.	.	.	.	E	.	.	.
(2) Plain	.	.	.	.	.	T	.	.	.
(3) Cipher	.	.	.	.	.	X	.	.	.

Here it will be noted that E_p in line (2) has a T_p on either side of it, at a distance of 13 intervals; the first encipherment (E_p by T_k) yields the same equivalent (X_c) as the second encipherment (T_p by E_k). Two cipher letters are here identical, at an interval equal to the length of the introductory key. But the converse is not true; that is, not *every* pair of identical letters in the cipher text represents a case of this type. For in this system identity in two cipher letters may be the result of the following three conditions each having a statistically ascertainable probability of occurrence:

(1) A given plain-text letter is enciphered by the same key letter two different times, at an interval which is purely accidental; the cipher equivalents are identical but could not be used to give any information about the length of the introductory key.

(2) Two different plain-text letters are enciphered by two different key letters; the cipher equivalents are fortuitously identical.

(3) A given plain-text letter is enciphered by a given key letter and later on the same plain-text letter serves to encipher another plain-text letter which is identical with the first key letter; the cipher equivalents are causally identical.

It can be proved that the probability for identities of the third type is greater than that for identities of either or both 1st and 2d types *for that interval which corresponds with the length of the introductory key*; that is, if a tabulation is made of the intervals between identical letters in such a system as the one being studied, the interval which occurs most frequently should coincide with the length of the introductory key. The demonstration of the mathematical basis for this fact is beyond the scope of the present text; but a practical demonstration will be convincing.

d. Let the illustrative message be transcribed in lines of say 11, 12, and 13 letters, as in Figure 10.

1	2	3	4	5	6	7	8	9	10	11
T	M	C	W	J	V	M	P	S	G	X
C	L	D	C	N	I	N	O	N	Y	G
U	O	I	N	P	E	T	X	Q	G	T
R	X	F	J	I	M	C	E	E	X	U
J	T	W	D	Y	X	A	Z	R	K	$\underline{G}$
V	A	M	X	K	F	O	D	W	N	$\underline{G}$
L	K	F	B	H	P	F	W	Q	Z	R
H	X	S	K	F	N	M	I	A	J	C
F	G	K	$\underline{P}$	Y	X	I	Y	M	P	R
X	E	O	$\underline{P}$	Q	W	W	R	V	C	W
J	S	E	$\underline{W}$	F	Z	M	C	L	O	P
I	U	G	$\underline{W}$	A	X	W	U	G	V	M
F	Y	X	J	X	W	Z	F	W	E	V
E	U	R	Z	R	H	H	G	U	T	Q
B	G									

1	2	3	4	5	6	7	8	9	10	11	12
T	M	$\underline{C}$	W	J	V	M	P	S	$\underline{G}$	X	C
L	D	$\underline{C}$	N	I	N	O	N	Y	$\underline{G}$	U	O
I	N	P	E	T	X	Q	G	T	R	X	F
J	I	M	C	E	E	X	U	J	T	W	D
Y	X	A	Z	R	K	G	V	A	M	X	K
$\underline{F}$	O	D	W	N	G	L	K	F	B	H	P
$\underline{F}$	W	Q	Z	R	H	X	S	K	F	N	M
I	A	J	C	F	G	K	P	Y	X	I	Y
M	P	R	X	$\underline{E}$	O	P	Q	W	W	R	V
C	W	J	S	$\underline{E}$	W	F	Z	M	C	L	O
P	I	U	G	W	A	X	W	U	G	$\underline{V}$	M
F	Y	X	J	X	W	Z	F	W	E	$\underline{V}$	E
U	R	Z	R	H	H	G	U	T	Q	B	G

b

1	2	3	4	5	6	7	8	9	10	11	12	13
T	M	C	W	J	V	M	P	S	G	X	C	L
D	C	N	I	N	O	N	Y	G	U	O	$\underline{I}$	N
P	$\underline{E}$	T	$\underline{X}$	Q	G	$\underline{T}$	R	X	F	J	$\underline{I}$	M
C	$\underline{E}$	E	X	U	J	$\underline{T}$	W	D	Y	X	A	Z
R	K	G	V	A	M	X	K	F	O	D	W	N
G	L	$\underline{K}$	$\underline{F}$	B	H	P	F	W	Q	Z	R	H
X	S	$\underline{K}$	$\underline{F}$	N	$\underline{M}$	I	A	J	C	F	G	K
P	Y	X	I	Y	$\underline{M}$	P	R	X	E	O	P	Q
W	W	R	V	C	W	J	S	E	W	F	Z	M
C	L	O	P	I	U	G	$\underline{W}$	A	X	$\underline{W}$	U	G
V	M	F	Y	X	J	X	$\underline{W}$	Z	F	$\underline{W}$	E	V
E	U	R	Z	R	H	H	G	U	T	Q	B	G

FIGURE 10.

In each transcription, every pair of superimposed letters is noted and the number of identities is indicated by ringing the letters involved, as shown above. The number of identities for an assumed introductory-key length 13 is 9, as against 3 for the assumption of a key of 11 letters, and 5 for the assumption of a key of 12 letters.

e. Once having found the length of the introductory key, two lines of attack are possible: the composition of the key may be studied, which will yield sufficient plain text to get a start toward solution; or, the message may be reduced to periodic terms and solved as a repeating-key cipher. The first line of attack will be discussed first, it being constantly borne in mind in this paragraph that the entire discussion is based upon the assumption that the cipher alphabets are known alphabets. The illustrative message of *b* above will be used.

5. Subsequent steps after determining the length of the introductory key.—*a*. Assume that the first letter of the introductory key is A and decipher the 1st cipher letter T_c (with direct standard alphabets). This yields T_p and the latter becomes the key letter for the 14th letter of the message. The 14th letter is deciphered: D_c $(T_k)=K_p$; the latter becomes the key letter for the 27th letter and so on, down the entire first column of the message as transcribed in lines of 13 letters. The same procedure is followed using B as the initial key letter, then C, and so on. The message as it appears for the first three trials (assuming A, B, then C as the initial key letter) is shown in Figure 11.

1	2	3	4	5	6	7	8	9	10	11	12	13
T	M	C	W	J	V	M	P	S	G	X	C	L
T												
D												
K												
P												
F												
C												
X												
R												
U												
G												
M												
X												
L												
P												
E												
W												
S												
C												
K												
V												
L												
E												
T												

(*a*) First column of Figure 10 (*c*) "deciphered" with initial θ_k=A.

1	2	3	4	5	6	7	8	9	10	11	12	13
T	M	C	W	J	V	M	P	S	G	X	C	L
S												
D												
L												
P												
E												
C												
Y												
R												
T												
G												
N												
X												
K												
P												
F												
W												
R												
C												
L												
V												
K												
E												
U												

(*b*) First column of Figure 10 (*c*) "deciphered" with initial θ_k=B.

1	2	3	4	5	6	7	8	9	10	11	12	13
T	M	C	W	J	V	M	P	S	G	X	C	L
R												
D												
M												
P												
D												
C												
Z												
R												
S												
G												
O												
X												
J												
P												
G												
W												
Q												
C												
M												
V												
J												
E												
V												

(*c*) First column of Figure 10 (*c*) "deciphered" with θ_k=C.

FIGURE 11.

b. Inspection of the results of these three trials soon shows that the entire series of 26 trials need not be made, for the results can be obtained from the very first trial. This may be shown graphically by superimposing merely the results of the first three trials *horizontally*. Thus:

		T	D	P	C	R	G	X	P	W	C	V	E
Cipher letters of Col. 1, Fig. 11		T	D	P	C	R	G	X	P	W	C	V	E
Keyletters	A	T	K	F	X	U	M	L	E	S	K	L	T
	B	S	L	E	Y	T	N	K	F	R	L	K	U
	C	R	M	D	Z	S	O	J	G	Q	M	J	V
	D	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓
	E												

FIGURE 12.

c. It will be noted that the vertical sequences in adjacent columns proceed in opposite directions, whereas those in alternate columns proceed in the same direction. The explanation for this alternation in progression is the same as in the previous case wherein this phenomenon was encountered (par. 2*b*), and the sequences in Figure 12 may now be completed very quickly. The diagram becomes as shown in Figure 13.

d. *One* of the horizontal lines or generatrices of figure 13 is the correct one; that is, it contains the actual plain-text equivalents of the 1st, 14th, 27th, . . . letters of the message. The correct generatrix can be selected by mere ocular examination, as is here possible (see generatrix marked by asterisk in Fig. 13), or it may be selected by a frequency test, assigning weights to each letter according to its normal plain-text frequency. (See par. 14*f* of *Military Cryptanalysis, Part II.*)

T	D	P	C	R	G	X	P	W	C	V	E
T	K	F	X	U	M	L	E	S	K	L	T
S	L	E	Y	T	N	K	F	R	L	K	U
R	M	D	Z	S	O	J	G	Q	M	J	V
Q	N	C	A	R	P	I	H	P	N	I	W
P	O	B	B	Q	Q	H	I	O	O	H	X
O	P	A	C	P	K	G	J	N	P	G	Y
N	Q	Z	D	O	S	F	K	M	Q	F	Z
M	R	Y	E	N	T	E	L	L	R	E	A*
L	S	X	F	M	U	D	M	K	S	D	B
K	T	W	G	L	V	C	N	J	T	C	C
J	U	V	H	K	W	B	O	I	U	B	D
I	V	U	I	J	X	A	P	H	V	A	E
H	W	T	J	I	Y	Z	Q	G	W	Z	F
G	X	S	K	H	Z	Y	R	F	X	Y	G
F	Y	R	L	G	A	X	S	E	Y	X	H
E	Z	Q	M	F	B	W	T	D	Z	W	I
D	A	P	N	E	C	V	U	C	A	V	J
C	B	O	O	D	D	U	V	B	B	U	K
B	C	N	P	C	E	T	W	A	C	T	L
A	D	M	Q	B	F	S	X	Z	D	S	M
Z	E	L	R	A	G	R	Y	Y	E	R	N
Y	F	K	S	Z	H	Q	Z	X	F	Q	O
X	G	J	T	Y	I	P	A	W	G	P	P
W	H	I	U	X	J	O	B	V	H	O	Q
V	I	H	V	W	K	N	C	U	I	N	R
U	J	G	W	V	L	M	D	T	J	M	S

FIGURE 13.

e. Identical procedure is followed with respect to columns 2, 3, 4, . . . of Figure 10*c*, with the result that the initial key word HORSECHESTNUT is reconstructed and the whole message may be now deciphered quite readily.

6. Conversion of foregoing aperiodic cipher into periodic form.—*a.* In paragraph 4 it was stated that an aperiodic cipher of the foregoing type may be reduced to periodic terms and solved as though it were a repeating-key cipher, provided the primary components are known sequences. The basis of the method lies in the phenomena noted in paragraph 2*b*. An example will be given.

b. Let the cipher text of the message of paragraph 4*b* be set down again, as in Figure 10*c*:

1	2	3	4	5	6	7	8	9	10	11	12	13
T	M	C	W	J	V	M	P	S	G	X	C	L
D	C	N	I	N	O	N	Y	G	U	O	I	N
P	E	T	X	Q	G	T	R	X	F	J	I	M
C	E	E	X	U	J	T	W	D	Y	X	A	Z
R	K	G	V	A	M	X	K	F	O	D	W	N
G	L	K	F	B	H	P	F	W	Q	Z	R	H
X	S	K	F	N	M	I	A	J	C	F	G	K
P	Y	X	I	Y	M	P	R	X	E	O	P	Q
W	W	R	V	C	W	J	S	E	W	F	Z	M
C	L	O	P	I	U	G	W	A	X	W	U	G
V	M	F	Y	X	J	X	W	Z	F	W	E	V
E	U	R	Z	R	H	H	G	U	T	Q	B	G

FIGURE 10c.

Using direct standard alphabets (Vigenère method), "decipher" the second line by means of the first line, that is, taking the letters of the second line as cipher text, those of the first line as key letters. Then use the thus-found "plain text" as "key letters" and "decipher" the third line of Figure 10*c*, as shown in Figure 14. Thus:

"Key"	T	M	C	W	J	V	M	P	S	G	X	C	L
Cipher	D	C	N	I	N	O	N	Y	G	U	O	I	N
"Plain"	K	Q	L	M	E	T	Z	J	O	O	R	G	C
"Key"	K	Q	L	M	E	T	Z	J	O	O	R	G	C
Cipher	P	E	T	X	Q	G	T	R	X	F	J	I	M
"Plain"	F	O	I	L	M	N	U	I	J	R	S	C	K

FIGURE 14.

Continue this operation for all the remaining lines of Figure 10*c* and write down the results in lines of 26 letters. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
T	M	C	$\underline{W}$	J	$\underline{V}$	M	P	S	G	X	$\underline{C}$	L	$\underline{K}$	$\underline{Q}$	L	$\underline{M}$	E	T	$\underline{Z}$	$\underline{J}$	O	O	R	G	C
F	O	I	L	M	N	U	I	J	R	S	$\underline{C}$	K	X	$\underline{Q}$	$\underline{W}$	$\underline{M}$	I	W	$\underline{Z}$	$\underline{O}$	U	H	F	Y	$\underline{P}$
U	U	$\underline{K}$	$\underline{J}$	S	Q	Y	W	L	H	Y	Y	Y	M	R	A	W	J	$\underline{R}$	R	$\underline{J}$	L	J	$\underline{B}$	T	J
$\underline{L}$	B	$\underline{K}$	$\underline{J}$	E	$\underline{V}$	R	R	Y	$\underline{T}$	E	N	B	E	X	N	Z	$\underline{U}$	$\underline{R}$	Y	A	Z	L	K	C	$\underline{P}$
S	Z	E	$\underline{W}$	I	F	L	$\underline{S}$	F	L	$\underline{V}$	X	X	$\underline{K}$	M	K	T	A	P	V	E	V	M	$\underline{B}$	X	J
$\underline{L}$	A	V	F	X	U	C	$\underline{S}$	E	$\underline{T}$	$\underline{V}$	H	M	T	U	$\underline{W}$	U	$\underline{U}$	N	F	$\underline{O}$	Q	A	V	U	U

FIGURE 15.

Now write down the real plain text of the message in lines of 26 letters. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
M	Y	L	E̲	F	T̲	F	L	A	N	K	I̲	S	R̲	E̲	C	E̲	I	V	I̲	N̲	G	H	E	A	V
Y	A	R	T	I	L	L	E	R	Y	F	I̲	R	E	E̲	N̲	E̲	M	Y	I̲	S̲	M	A	S	S	I̲
N	G	T̲	R̲	O	O	P	S	T	O	L	E	F	T	F	R	O	N	T̲	A	N̲	D	C	O̲	N	C
E̲	N	T̲	R̲	A	T̲	I	N	G	A̲	R	T	I	L	L	E	R	Y̲	T̲	H	E	R	E	X	W	I̲
L	L	N	E̲	E	D	C	O̲	N	S	I̲	D	E	R̲	A	B	L	E	R	E	I	N	F	O̲	R	C
E̲	M	E	N	T	S	T	O̲	M	A̲	I̲	N	T	A	I	N̲	M	Y̲	P	O	S̲	I	T	I	O	N

FIGURE 16.

c. When the underlined repetitions in Figures 15 and 16 are compared, they are found to be identical in the respective columns, and if the columns of Figure 15 are tested, they will be found to be monoalphabetic. The cipher message now gives every indication of being a repeating-key cipher. It is not difficult to explain this phenomenon in the light of the demonstration given in paragraph 3*g*. First, let the key word HORSECHESTNUT be enciphered by the following alphabet:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B

"Plain" .... H O R S E C H E S T N U T
"Cipher" .... T M J I W Y T W I H N G H

Then let the message MY LEFT FLANK, etc., be enciphered by direct standard alphabets as before, but for the key add the monoalphabetic equivalents of HORSECHESTNUT TMJIW. . . to the key itself, that is, use the 26-letter key HORSECHESTNUTTMJIWYTWIHNGH in a repeating-key manner. Thus (Fig. 17):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Key..................	H	O	R	S	E	C	H	E	S	T	N	U	T	T	M	J	I	W	Y	T	W	I	H	N	G	H
Plain..................	M	Y	L	E	F	T	F	L	A	N	K	I	S	R	E	C	E	I	V	I	N	G	H	E	A	V
Cipher................	T	M	C	W	J	V	M	P	S	G	X	C	L	K	Q	L	M	E	T	Z	J	O	O	R	G	C
Plain..................	Y	A	R	T	I	L	L	E	R	Y	F	I	R	E	E	N	E	M	Y	I	S	M	A	S	S	I
Cipher................	F	O	I	L	M	N	U	I	J	R	S	C	K	X	Q	W	M	I	W	Z	O	U	H	F	Y	P
Plain..................	N	G	T	R	O	O	P	S	T	O	L	E	F	T	F	R	O	N	T	A	N	D	C	O	N	C
Cipher................	U	U	K	J	S	Q	Y	W	L	H	Y	Y	Y	M	R	A	W	J	R	R	J	L	J	B	T	J
Plain..................	E	N	T	R	A	T	I	N	G	A	R	T	I	L	L	E	R	Y	T	H	E	R	E	X	W	I
Cipher................	L	B	K	J	E	V	R	R	Y	T	E	N	B	E	X	N	Z	U	R	Y	A	Z	L	K	C	P
Plain..................	L	L	N	E	E	D	C	O	N	S	I	D	E	R	A	B	L	E	R	E	I	N	F	O	R	C
Cipher................	S	Z	E	W	I	F	L	S	F	L	V	X	X	K	M	K	T	A	P	V	E	V	M	B	X	J
Plain..................	E	M	E	N	T	S	T	O	M	A	I	N	T	A	I	N	M	Y	P	O	S	I	T	I	O	N
Cipher................	L	A	V	F	X	U	C	S	E	T	V	H	M	T	U	W	U	U	N	F	O	Q	A	V	U	U

FIGURE 17.

The cipher resultants of this process of *enciphering* a message coincide exactly with those obtained from the "*deciphering*" operation that gave rise to Figure 15. How does this happen?

d. First, let it be noted that the sequence TMJI . . . , which forms the second half of the key for enciphering the text in Figure 17 may be described as the standard alphabet *complement* of the sequence HORSECHESTNUT, which forms the first half of that key. Arithmetically, the sum of a letter of the first half and its homologous letter in the second half is 26. Thus:

$$
\begin{aligned}
H+T&=\ 7+19=26=0\\
O+M&=14+12=26=0\\
R+J&=17+\ 9=26=0\\
S+I&=18+\ 8=26=0\\
E+W&=\ 4+22=26=0
\end{aligned}
$$

That is, every letter of HORSECHESTNUT plus its homologous letter of the sequence TMJIWYTYIHNGH equals 26, which is here the same as zero. In other words, the sequence TMJIWYTWIHNGH is, by cryptographic arithmetic, equivalent to "minus HORSECHESTNUT." Therefore, in Figure 17, *enciphering* the second half of each line by the key letters TMJIWYTWIHNCH (i. e., adding 19, 12, 9, 8, . . .) is the same as *deciphering* by the key letters HORSECHESTNUT (i. e., subtracting 7, 14, 17, 18, . . .). For example:

$R_p(T_k)=17+19=36=10=K$, and
$R_p(-H_k)=17-7=10=K$

$E_p(M_k)=4+12=16=Q_c$, and
$E_p(-O_k)=4-14=(26+4)-14=16=Q_c$, and so on.

e. Refer now to Figure 15. The letters in the first half of line 1, beginning TMCWJ . . . are identical with those in the first half of line 1 of Figure 17. They must be identical because they are produced from identical elements. The letters in the second half of this same line in Figure 15, beginning KQLME . . . were produced by *deciphering* the letters in the second line of Figure 10*c*. Thus (taking for illustrative purposes only the first five letters in each case):

KQLME = DCNIN − TMCWJ
But DCNIN = RECEI + MYLEF
And TMCWJ = MYLEF + HORSE
Hence, KQLME = (RECEI + MYLEF) − (MYLEF + HORSE)
Or, KQLME = RECEI − HORSE (1)

As for the letters in the second half of line 1 of Figure 17, also beginning KQLME . . . , these letters were the result of *enciphering* RECEI by TMJIW.
Thus:

KQLME = RECEI + TMJIW

But it has been shown in subparagraph *d* above that

TMJIW = − HORSE

Hence, KQLME = RECEI + (− HORSE)
Or, KQLME = RECEI − HORSE (2)

Thus, equations (1) and (2) turn out to be identical but from what appear to be quite diverse sources.

f. What has been demonstrated in connection with the letters in line 1 of Figures 15 and 17 holds true for the letters in the other lines of these two figures, and it is not necessary to repeat the explanation. The steps show that the originally aperiodic, auto-key cipher has been converted, through a knowledge of the primary components, into a repeating-key cipher with a period twice the length of the introductory key. The message may now be solved as an ordinary repeating-key cipher.

g. (1) The foregoing case is based upon encipherment by the enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{c/2}$. When encipherment by the enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/2}=\Theta_{c/1}$ has been followed, the conversion of a plain-text auto-keyed cipher yields a repeating-key cipher with a period equal to the length of the introductory key. In this conversion, the enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{c/2}$ are used in finding equivalents.

(2) An example may be useful. Note the encipherment of the following message by auto-key method by enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/2}=\Theta_{c/1}$.

```
T U E S D A Y|I N F O R M A T I O N F R O M R E L I A B L E S O U R C E S I N D I C
I N F O R M A T I O N F R O M R E L I A B L E S O U R C E S I N D I C A T E S T H E
P T B W O M C L V J Z O F O T J Q Y D J N Z N O D M R B T O Q Z J R A W B W F Q Z C
```

(3) If the message is written out in lines corresponding to the length of the introductory key, and each line is *enciphered* by the one directly above it, using the enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{c/2}$ in finding equivalents, the results are as shown in Figure 22*b*. But if the same message is *enciphered* by equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/2}=\Theta_{c/1}$, using the word TUESDAY as a repeating key, the cipher text (Fig. 18*c*) is identical with that obtained in Figure 18*b* by enciphering each successive line with the line above it.

Original cipher text		Original cipher text and converted text		Repeating key encipherment
				T U E S D A Y
				I N F O R M A
P T B W O M C	⟷	P T B W O M C	⟷	P T B W O M C
L V J Z O F O	⟷	L V J Z O F O		
				T I O N F R O
		A O K V C R Q	⟷	A O K V C R Q
T J Q Y D J N	⟷	T J Q Y D J N		
				M R E L I A B
		T X A T F A D	⟷	T X A T F A D
Z N O D M R B	⟷	Z N O D M R B		
				L E S O U R C
		S K O W R R E	⟷	S K O W R R E
T O Q Z J R A	⟷	T O Q Z J R A		
				E S I N D I C
		L Y E V A I E	⟷	L Y E V A I E
W B W F Q Z C	⟷	W B W F Q Z C		
				A T E S T H E
		H Z A A Q H G	⟷	H Z A A Q H G
a		*b*		*c*

FIGURE 18.

(4) Now note that the sequences joined by arrows in Figure 18 *b* and *c* are identical and since it is certain that Figure 18*c* is periodic in form because it was enciphered by the repeating-key method, it follows that Figure 18*b* is now also in periodic form, and in that form the message could be solved as though it were a repeating-key cipher.

h (1) In case of primary components consisting of a direct normal sequence sliding against a reversed normal (U. S. Army disk), the process of converting the auto-key text to periodic terms is accomplished by using two direct normal sequences and "deciphering" each line of the text (as transcribed in periods) by the line above it. For example, here is a message auto-enciphered by the aforementioned disk, with the initial key word TUESDAY:

```
TUESDAY|INFORMATIONFROMRELIABLESOURCESINDIC
INFORMATIONFROMRELIABLESOURCESINDICATESTHE
LHZEMOYPFRBMVMHRKCXRNBNMXOJZHMKBRJAEZEVKBY
```

(2) The cipher text is transcribed in periods equal to the length of the initial key word (7 letters) and the 2d line is "deciphered" with key letters of the 1st line, using enciphering equations $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/1}=\theta_{c/2}$. The resultant letters are then used as key letters to "decipher" the 3d line of text and so on. The results are as seen in Figure 19*b*. Now let the original message be *enciphered* in repeating-key manner by the disk, with the key word TUESDAY, and the result is Figure 19*c*. Note that the odd or alternate lines of Figure 19*b* and *c* are identical, showing that the auto-key text has been converted into repeating-key text.

Original cipher text		Original cipher text and converted text		Repeating key encipherment
				T U E S D A Y
				I N F O R M A
L H Z E M O Y	→	L H Z E M O Y	→	L H Z E M O Y
P F R B M V M	→	P F R B M V M		T I O N F R O
		A M Q F Y J K	→	A M Q F Y J K
H R K C X R N	→	H R K C X R N		M R E L I A B
		H D A H V A X	→	H D A H V A X
B N M X O J Z	→	B N M X O J Z		L E S O U R C
		I Q M E J J W	→	I Q M E J J W
H M K B R J A	→	H M K B R J A		E S I N D I C
		P C W F A S W	→	P C W F A S W
E Z E V K B Y	→	E Z E V K B Y		A T E S T H E
		T B A A K T U	→	T B A A K T U
a		*b*		*c*

FIGURE 19.

i. The foregoing procedures indicate a simple method of solving ciphers of the foregoing types, when the primary components or the secondary cipher alphabets are known. It consists in assuming introductory keys of various lengths, converting the cipher text into repeating-key form, and then examining the resulting diagrams for repetitions. When a correct key length is assumed, repetitions will be as numerous as should be expected in ciphers of the repeating-key class; incorrect assumptions for key length will not show so many repetitions.

j. All the foregoing presupposes a knowledge of the cipher alphabets involved. When these are unknown, recourse must be had to first principles and the messages must be solved purely upon the basis of probable words, and repetitions, as outlined in paragraphs 27–28.

INDEX

Page

Aperiodic systems 1
Arithmetical equivalent of normal sliding-strip encipherment 100
Auto-key encipherment, two basic methods of 28
Auto-key systems:
  Solution of 28, 98
  Characteristics of 31
  Concluding remarks on 48
Auto keying:
  Cipher text 28, 30
  Plain text 28, 45, 98
Avoiding periodicity, methods of 1

Base letter 32, 36
Basic period masked by aperiodic repetitions 16
Blocking out isomorphs 14
Blocking out words 14
Book as key 5, 50, 51, 53, 74

Chi test 73, 89
  Applying the 77, 79
  Derivation of 75
  Example of application of 79
  In matching shifted distributions 77
  Nature of 74
Coincidence test 58
  Application of 63
  Basic theory of 58
Coincidence, intermittent 18
Combining individual frequency distributions 74
Comparisons for coincidence 60
Constant-length, plain-text groupings 2, 19
Continuous-key system 50, 51
Conversion of an aperiodic cipher into periodic form 112
Converting auto-key text to periodic terms 112
Cross-product sum or χ test 73
Cryptanalytic coincidence test 58
Cryptographic arithmetic 100
Cryptographic periodicity, nature of 1
Cyclic phenomena 1

Enciphering equations 7
Encipherment by word lengths 5
Extended keys 50

Page

Formulae, idiomorphic 8
Frequency distribution square 81

General solution for ciphers involving a long keying sequence 56
Groupings:
  Constant-length plain-text 2, 19
  Irregular 16
  Variable-length, plain-text 5

Identity or coincidence 58
Idiomorphism 8
Indicators 56
Influence letter 21
Initial key 28, 32
Interlocking messages by repetitions 56
Intermittent coincidences 18
Interrupting a cyclic keying sequence 19
Interrupting the key, three basic methods of 19
Interruptions, keying 19
Interruptor 21
  Cipher-text letter as 25
  Plain-text letter as 21
  Disadvantages of 27
Introductory key 28, 32
  Consisting of more than one letter 32, 45, 107
Irregular interruptions in keying sequence 19
Isomorphic sequences 11
Isomorphism:
  Detection of 11
  Illustration of the use of 11, 39
  Phenomena of 11
Isomorphs, blocking out of 14

Kappa test 58
  Application of 63
Keying, fixed 1
Keying cycles, interaction of 4
Keying units:
  Constant length 5
  Variable length 19
Keys, extended; nonrepeating; running 50

Lengthening keys 50

Page

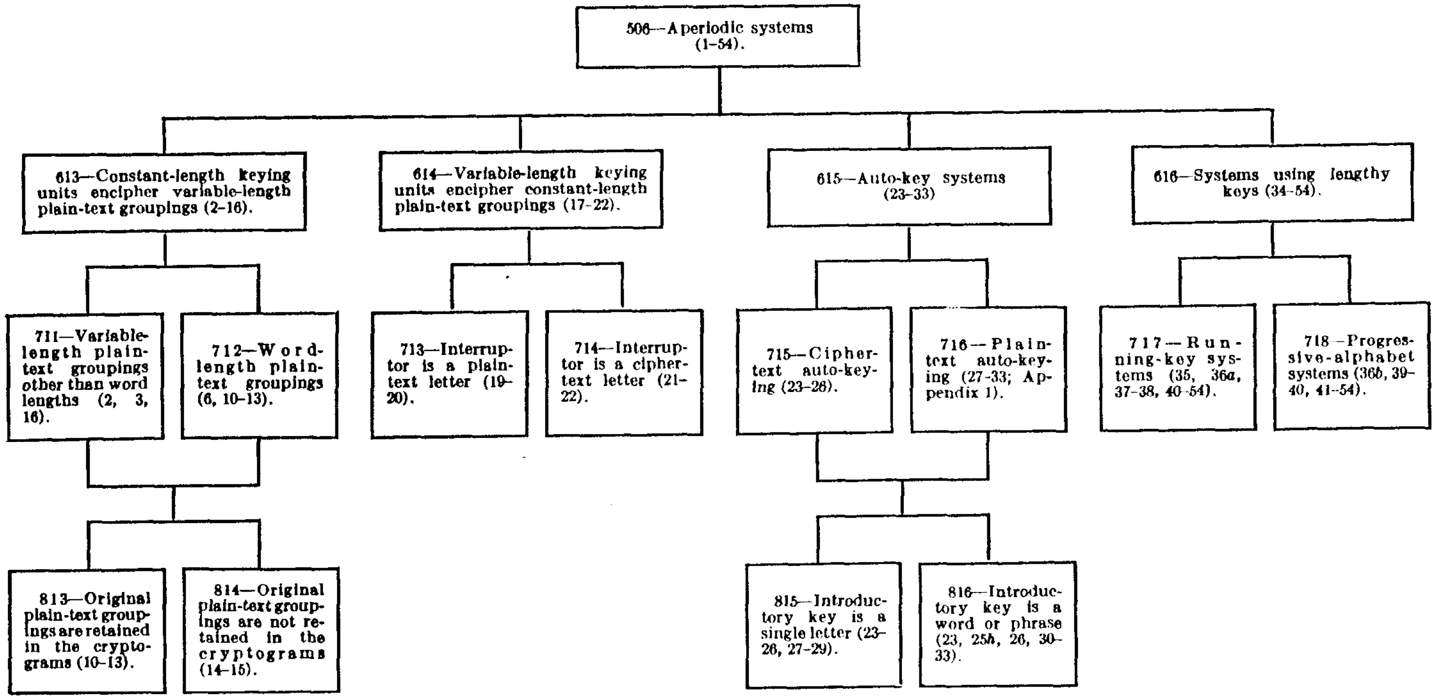
- Lengthy keys:
  - Systems using 1, 50
  - Mechanical methods of producing 52
- Making the κ test, general procedure 62
- Matching of frequency distributions 73
- Monoalphabeticity or Φ test 94
- Monographic coincidence, probability of 58
- Nonrepeating key system 50
- Overlap 51
- Partial periodicity 3
- Patterns:
  - Idiomorphic 8
  - Word 8
- Period, apparent; basic; complete; hidden; latent; patent; primary; resultant; secondary 4
- Periodicity, masked 16
- Periods:
  - Component 4
  - Superimposed 23
- Phi test 94
  - Applying 95
  - Derivation of 93
  - Purpose of 93
  - Related to χ test 96
- Probability, theory of 58
- Probability of monographic coincidence 58
- Progressive-alphabet cipher, solution of 52, 55, 82
- Progressive alphabet system 52, 55
- Reconstruction skeleton 9
- Repetitions:
  - Completely periodic 3

Page

- Repetitions—Continued.
  - Nonperiodic 3
  - Partially periodic 3
  - Significant 25
- Resultant key 4, 52
- Running-key cipher, solution of 53, 56, 63, 71
- Running-key system 51
- Secondary key 4, 52
- Separators, word 15
- Sequences, uninterrupted 26
- Solution by superimposition 23, 53, 58
- Spurious plain text 43, 104
- Statistical test 26
- Stereotypic phraseology 14
- Superimposable periods 23
- Superimposed sequences and the coincidence test 58
- Superimposition 53
  - Basic principles of 53
  - Correct and incorrect 58
  - Solution by 53
  - Diagram 61
- Synoptic table 97, 119
- Symmetry of position, direct 9
- Variable-length:
  - Groupings of keying sequence 19
  - Key enciphering 19
  - Plain-text groupings 5
- Vigenère method 46
- Wheatstone cryptograph 52
- Word habits of the enemy, familiarity with 14
- Word-length encipherment, solution of 5
- Word separators 15

Analytical Key for Military Cryptanalysis, Part III

[Numbers in parentheses refer to Paragraph Numbers in this text]


506—Aperiodic systems (1-54).
613—Constant-length keying units encipher variable-length plain-text groupings (2-16).
614—Variable-length keying units encipher constant-length plain-text groupings (17-22).
615—Auto-key systems (23-33)
616—Systems using lengthy keys (34-54).
711—Variable-length plain-text groupings other than word lengths (2, 3, 16).
712—Word-length plain-text groupings (6, 10-13).
713—Interruptor is a plain-text letter (19-20).
714—Interruptor is a cipher-text letter (21-22).
715—Cipher-text auto-keying (23-26).
716—Plain-text auto-keying (27-33; Appendix 1).
717—Running-key systems (35, 36a, 37-38, 40-54).
718—Progressive-alphabet systems (36b, 39-40, 41-54).
813—Original plain-text groupings are retained in the cryptograms (10-13).
814—Original plain-text groupings are not retained in the cryptograms (14-15).
815—Introductory key is a single letter (23-26, 27-29).
816—Introductory key is a word or phrase (23, 25b, 26, 30-33).

BOOKS IN THE CRYPTOGRAPHIC SERIES

I - Manual for the Solution of Military Ciphers, Parker Hitt
2 - Cryptanalysis of the Simple Substitution Cipher with Word Divisions, Wayne G. Barker
3 - Elements of Cryptanalysis, William F. Friedman
4 - Statistical Methods in Cryptanalysis, Solomon Kullback, Ph.D.
5 - Cryptography and Cryptanalysis Articles, Volume 1, edited by William F. Friedman
6 - Cryptography and Cryptanalysis Articles, Volume 2, edited by William F. Friedman
7 - Elementary Military Cryptography, William F. Friedman
8 - Advanced Military Cryptography, William F. Friedman
9 - War Secrets in the Ether, Volume 1 [Parts I **and** II], Wilhelm F. Flicke
10 **- War Secrets in the Ether, Volume 2 [Part** III], Wilhelm F. Flicke
11 - Solving German Codes in World War I, William F. Friedman
12 - History of the Use **of Codes,** William F. Friedman
13 - The Zimmermann Telegram of **January 16, 1917 and its Cryptographic Background,** William F. Friedman and Charles J. Mendelsohn, Ph.D.
14 - Manual of Cryptography, Luigi Sacco
16 - The Origin and Development of the Army Security Agency, 1917-1947
17 - **Cryptanalysis of the Hagelin Cryptograph,** Wayne G. Barker
18 **- The Contribution of the Cryptographic Bureaus in the World War, Yves** Gylden
19 - Course in Cryptography, Marcel Givierge
20 - History of Codes and Ciphers in the United States Prior to World War I
21 - History of Codes and Ciphers in the United States During World War I
22 - History of Codes and Ciphers in the United States During the Period Between-the World Wars, Part I. 1919-1929
23 - The Riverbank Publications, Volume 1, William F. Friedman
24 - The Riverbank Publications, Volume 2, William F. Friedman
25 - The Riverbank Publications, Volume 3, William F. Friedman
26 - Cryptanalysis of an Enciphered Code Problem - Where an “Additive” Method of Encipherment has been Used, Wayne G. Barker
27 - The Voynich Manuscript - An Elegant Enigma, M.E. D’lmperio
28 - Manual of Cryptography, British War Office
30’ - Military Cryptanalysis, Part I, William F. Friedman
31 - Speech and Facsimile Scrambling and Decoding - A Basic Text On Speech Scrambling
32 - Computer Simulation for Classical Substitution Cryptographic Systems, Rudolph F. Lauer
33 - Course In Cryptanalysis, Volume I. (Explanatory Text and Short -Exercises), British War Office
34 - Course in Cryptanalysis, Volume 11. (Figures and Cipher Texts), British War Office
35 - The Origin and Development of The National Security Agency, George **A.** Brownell
36 **- Treatise on Cryptography,** Andre Lange and E.A. Soudart
37 - **Solving Cipher Secrets,** M.E, Ohaver
38 - Cryptography, Andre Langie
39 - Cryptanalysis of **Shift-Register Generated Stream Cipher Systems,** Wayne G. Barker
40 - Military Cryptanalysis, Part II, William F. Friedman
41 - Elementary Course in Probability for the Cryptanalyst [Revised Edition], Andrew M. Gleason
42 - Military Cryptanalytics, Part I, Volume 1, William F. Friedman and Lambros D. Callimahos
43 - Military Cryptanalytics, Part I, Volume 2, William F. Friedman and Lambros D. Callimahos
44 - Military Cryptanalytics, Part II, Volume 1, Lambros D. Callimahos and William F. Friedman
45 - Military Cryptanalytics, Part II, Volume 2, Lambros D. Callimahos and William F. Friedman

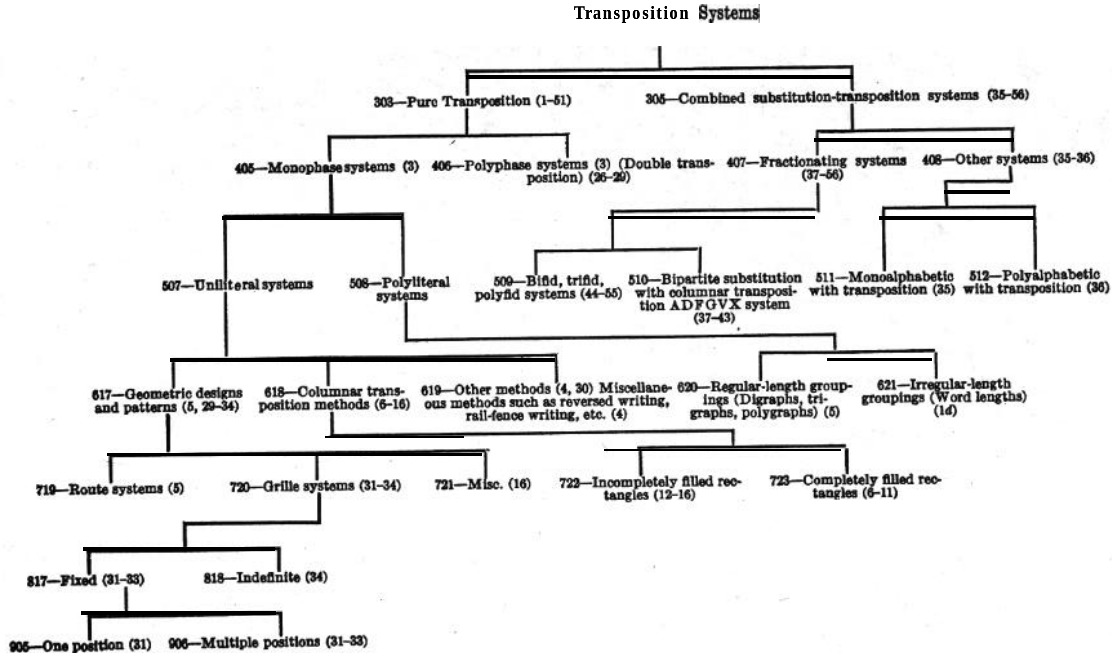
AEGEAN PARK PRESS, P. O. Box 2837, Laguna Hills, California 92654-0837

A CRYPTOGRAPHIC SERIES

61

MILITARY CRYPTANALYSIS
PART IV

Transposition and Fractionating Systems


Transposition Systems
303—Pure Transposition (1–51)
305—Combined substitution-transposition systems (35–56)
405—Monophase systems (3)
406—Polyphase systems (3) (Double transposition) (26–29)
407—Fractionating systems (37–56)
408—Other systems (35–36)
507—Uniliteral systems
508—Polyliteral systems
509—Bifid, trifid, polyfid systems (44–55)
510—Bipartite substitution with columnar transposition ADFGVX system (37–43)
511—Monoalphabetic with transposition (35)
512—Polyalphabetic with transposition (36)
617—Geometric designs and patterns (5, 29–34)
618—Columnar transposition methods (6–16)
619—Other methods (4, 30) Miscellaneous methods such as reversed writing, rail-fence writing, etc. (4)
620—Regular-length groupings (Digraphs, trigraphs, polygraphs) (5)
621—Irregular-length groupings (Word lengths) (1d)
719—Route systems (5)
720—Grille systems (31–34)
721—Misc. (16)
722—Incompletely filled rectangles (12–16)
723—Completely filled rectangles (6–11)
817—Fixed (31–33)
818—Indefinite (34)
905—One position (31)
906—Multiple positions (31–33)

(Numbers in parenthesis refer to paragraph numbers in this text.)

by

William F. Friedman

From Aegean Park Press

MILITARY CRYPTANALYSIS PART IV **1**
Transposition and Fractionating Systems 1

SECTION I **6**
GENERAL 6

SECTION II **8**
SOLUTION OF SIMPLE TRANSPOSITION CIPHERS ... 8

SECTION III **23**
INCOMPLETELY-FILLED RECTANGLES 23

SECTION IV **42**
OPPORTUNITIES AFFORDED BY STUDYING.............. 42

SECTION V **45**
SPECIAL SOLUTIONS FOR TRANSPOSITION 45

SECTION VI **85**
PRINCIPLES OF MATRIX RECONSTRUCTION 85

SECTION VII **91**
SOLUTION OF GRILLES.. 91

SECTION VIII **100**
COMBINED SUBSTITUTIOX-TRANSPOSITION 100

SECTION IX **103**
SOLUTION OF THE ADFGVX SYSTEM 103

SECTION X **150**
SOLUTION OF BIFID FRACTIONATING SYSTEMS 150

ANALYTICAL KEY FOR MILITARY CRYPTANALYSIS, ... **198**

INDEX **199**

MILITARY CRYPTANALYSIS

Part IV

TRANSPOSITION AND FRACTIONATING SYSTEMS

BY

WILLIAM F. FRIEDMAN

This is a quality reproduction of a U.S. Military text, originally published in 1941 - declassified from CONFIDENTIAL in December 1992.

ISBN: O-89412-198-7 **(soft cover)**

AEGEAN PARK PRESS
P.O. Box 2837
Laguna Hills, California 92654
(714) 586-8811
FAX (714) 586-8269

Manufactured in the United States of America

The Golden Guess
Is Morning-Star to the full round of Truth.
-Tennyson.

MILITARY CRYPTANALYSIS, PART IV.

TRANSPOSITION AND FRACTIONATING SYSTEMS

CONTENTS

Section		Paragraphs	Pages
I.	General	1-3	I-2
II.	Solution of simple transposition ciphers	4-11	3-17
III.	Incompletely-filled rectangles	12-16	18-36
IV.	Opportunities afforded by studying errors and blunders made by enemy cryptographers	17-19	37-39
V.	Special solutions for transposition ciphers	20-29	40-79
VI.	Principles of matrix reconstruction	30-31	8O-84
VII.	Solution of grilles	32-34	85-93
VIII.	Combined substitution-transposition systems	35-36	94-96
IX.	Solution of the ADFGVX system	37-43	97-143
X.	Solution of bifid fractionating systems	44-56	144-184
XI.	Analytical key	57	185
	Index	186-188	

SECTION I

GENERAL

Paragraph
Introductory remarks concerning transposition ciphers 1
Basic mechanism of transposition ciphers 2
Monophase and polyphase transposition 3

1. **Introductory remarks concerning transposition ciphers.**—*a.* As stated in a previous text, transposition ciphers are roughly analogous to "jigsaw puzzles" in that all the pieces of which the original is composed are present but are merely disarranged. The pieces into which the picture forming the basis of a jigsaw puzzle may be divided are usually quite irregular in size and shape; the greater the amount of irregularity, as a rule, the greater the difficulty in reassembling the pieces in proper order. In this respect, too, transposition ciphers are analogous to jigsaw puzzles, for the greater the amount of distortion to which the plain text is subjected in the transposition process, the more difficult becomes the solution.

b. In jigsaw puzzles there is usually no regularity about the size of the individual pieces into which the original picture has been cut, and this feature, of course, materially contributes to the difficulty in reconstructing the picture. There are, to be sure, limits (dictated by considerations of practicability) which serve to prevent the pieces being made too small, for then they would become unmanageable; on the other hand, there are also limits which must be observed in respect to the upper magnitude of the pieces, for if they are made too large the puzzle becomes too easy to solve. These features of jigsaw puzzles also have their analogies in transposition methods. In the latter, if the textual units to be subjected to transposition are made quite large, say entire sentences, the difficulties a cryptanalyst will have in reconstructing the text are practically nil; on the other hand, if these textual units are made quite small, even smaller than single letters,[1] then the reconstruction of the transposition text by a cryptanalyst often becomes a very difficult matter. In between these two extremes there may be various degrees of fragmentation, limited only by considerations of practicability.

c. It is fortunate, however, that the cryptanalyst does not, as a rule, have to contend with problems in which the size of the textual units varies within the same message, as is the case in jigsaw puzzles. It is perhaps possible to devise a transposition system in which the text is divided up in such a manner that entire sentences, whole words, syllables, individual letters, and fractions of letters form the units for transposition; but it is not difficult to imagine how impractical such a scheme would be for regular communication, and it may be taken for granted that such irregularity in size of textual units will not be encountered in practical communication.

d. The days when the simple methods of word transposition were sufficient for military purposes have long since passed by, and it is hardly to be expected that cryptograms of such ineffectual nature will be encountered in the military communications of even the smaller armies of today. However, in time of emergency, when a counter-espionage censorship is exercised over internal communications, it is possible that isolated instances of simple word transposition may be encountered. The solution of such cases should present no difficulties, unless numerous code names and nulls are also used in the cryptograms. Mere experimentation with the cryptograms, trying various types and sizes of rectangles, will usually disclose the secret text. If code names are used and the context gives no clue to the identity of the persons or places mentioned, it may be necessary to wait until additional messages become available, or, lacking such a possibility, there is usually sufficient justification, under the exigencies of war, to compel the correspondents to reveal the meaning of these code names.

---

[1] Reference is here made to so-called fractionating systems. See Special Text No. 166, *Advanced Military Cryptography*, sec. XI.

e. Although transposition ciphers, as a general rule, are much less complex in their mechanics than are substitution ciphers, the cryptanalyst usually experiences a feeling of distaste and dismay when confronted with unknown ciphers of this category. There are several reasons for his dislike for them. In the first place, although transposition ciphers are admittedly less intricate than substitution ciphers, as a general rule there are not nearly so many cryptanalytic tools and "tricks" to be used in the solution of the former as there are in the latter, and therefore the mental stimulus and satisfaction which the cryptanalyst usually derives and regards as part of the reward for his hard labor in solving a cipher is often missing in the case of transposition ciphers. In the second place, despite their lack of complexity, the solution of transposition ciphers often involves a tremendous amount of time and labor most of which commonly turns out to be fruitless experimentation. Thirdly, in modern military communication, transposition methods are usually not employed alone but in conjunction with substitution methods—and then the problems may become difficult indeed, for usually before the substitution can be attacked it is necessary first to uncover the transposition. Finally, in working with transposition ciphers a much higher degree of accuracy in mere mechanical operations is required than in working with substitution ciphers, because the accidental omission or addition of a single letter will usually necessitate rewriting the work sheets applying to entire messages and starting afresh. Thus, this sort of work calls for a constant state of concentrated attention, with its resulting state of psychological tension, which takes its toll in mental wear and tear.

2. Basic mechanism of transposition ciphers.—*a.* Basically, all transposition ciphers involve at least two processes: (1) Writing the plain-text units (usually single letters) within a specific regular or irregular two-dimensional design called a "matrix," "cage," "frame," "rectangle," *etc.*, in such a prearranged manner that the said units are distributed regularly or irregularly throughout the various cells or subsections of that design; (2) removing the plain-text units from the design in such a prearranged manner as to change the original sequence in which they followed one another in the plain text, thus producing cipher text. Since the first process consists of inscribing the text within the design, it is technically referred to as the process of *inscription;* and since the second process consists of transcribing the text from the design, it is technically referred to as that of *transcription.* Either or both processes may be repetitive, by prearrangement of course, in which case the intermediate steps may be referred to as processes of *rescription,* or *rescriptive* processes.

b. It is hardly necessary at this point to give the student any indications as to how to differentiate a transposition from a substitution cipher. If a review is necessary, however, he is referred to Section IV of Military Cryptanalysis, Part I.

3. Monophase and polyphase transposition.—*a.* As may be inferred from the foregoing definitions, when a transposition system involves but a single process of inscription, followed by a single process of transcription, the system may be referred to as *monophase transposition,* commonly called *single transposition.* When one or more rescriptive processes intervene between the original inscription and the final transcription the system may be referred to as *polyphase transposition.* As a general rule, the solution of the latter type is much more difficult than the former, especially when the successive transpositions are theoretically correct in principle.

b. Any system which is suited for monophase transposition is also usually suited for polyphase transposition, the processes of inscription, rescription and transcription being accomplished with the same or with different keys.

Section II

SOLUTION OF SIMPLE TRANSPOSITION CIPHERS

Paragraph
Simple types of transposition 4
The principles of solution of uniliteral route-transposition ciphers 5
Keyed columnar transposition with completely-filled rectangles 6
Example of solution 7
The probable-word method of solution 8
General remarks on solution 9
Reconstruction of literal key 10
Column and row transposition 11

4. Simple types of transposition.—*a.* The simple cases of reversed writing, vertical writing, or rail fence writing hardly require serious attention, since they may be solved almost by inspection. These methods are included here only because they may be encountered in censorship operations.

b. The low degree of cryptographic security afforded by these methods may be increased to a slight degree by adding nulls or by disguising the original word lengths, and regrouping into false words or into groups of regular length.

c. Some examples of these simplest types of transposition follow. Let the message be: BRIDGE DESTROYED AT ELEVEN PM.

(1) Reversing only the words and retaining original word lengths:
Cipher... EGDIRB DEYORTSED TA NEVELE MP

(2) Reversing only the words and regrouping into false word lengths:
Cipher... EGDIRB DEYORT SEDTA NEVE LEMP

(3) Reversing the whole text and regrouping into fives:
Cipher... MPNEV ELETA DEYOR TSEDE GDIRB

(4) Reversing the whole text, regrouping into fives, and inserting a null in every fifth position:
Cipher... TRIMM PNEVP ELETA ADEYR ORTSL
EDEGU DIRBM

(5) Writing the text vertically in two columns and taking the resulting digraphs for the cipher text, as shown at the side. The cipher message becomes:

```
B S    B R
R T    I D
I R    G E
D O    D E
G Y    S T
E E    R O
D D    Y E
E      D
```

BSRTI RDOGY EEDDE, *or*
BIGDS RYDRD EETOE

These simple types can be solved merely by inspection.

5. The principles of solution of uniliteral route-transposition ciphers.—*a.* The so-called uniliteral route-transposition methods are next to be examined. The solution of cryptograms enciphered by these methods is a matter of experimenting with geometric figures, usually rectangles, of various dimensions suggested by the total number of letters in the message, then inspecting these rectangles, searching for whole words or the fragments of words by reading horizontally, diagonally, vertically, spirally, and so on.[1] (See Special Text No. 165, *Elementary Military Cryptography*, 1935, pars. 20, 21.)

b. The amount of experimentation that must be performed in the solution of ciphers of this type may be materially shortened by means of formulae and tables constructed for the purpose. But because ciphers of this type are of infrequent occurrence today, these formulae and tables are only occasionally useful and hence they have not been included in this text.[2]

6. Keyed columnar transposition with completely-filled rectangles.—*a*. In practical cryptography, the dimensions of the transposition rectangle, as a general rule, cannot vary between large limits; that is, it can be assumed in practice that rectangles based upon lines of writing containing less than 5 letters or more than 25 letters will not commonly be encountered. If the width, that is, the number of columns, is determined by a key, then the number of rows becomes a function of the length of the message to be enciphered. If the latter is very long, longer than can be conveniently handled without too many errors, it is a common practice to break up a message into two or more parts and treat each part as though it were a separate communication. Such parts are commonly termed *sections*.

b. When the last row of a transposition rectangle is completely filled, the solution of the resulting cryptogram is considerably more simple than when this is not the case.[3] Consequently, this will the constitute first case to be studied.

---

[1] It is interesting to observe that Daniel, of Biblical fame, was apparently the first cryptanalyst in history (as well as one of the earliest interpreters of dreams), for he solved the cryptogram in the "handwriting on the wall," obtaining as his decipherment words which he interpreted as predicting the downfall of Belshazzar and his dynasty (Daniel V: 1–28). The following partial account of the episode is not as enlightening as one might wish, but it is probably the best explanation available. It is taken from Dr. Max Seligsohn's article on the subject in *The Jewish Encyclopedia*, vol. 8, pp. 490–491 (1925): "MENE, MENE, TEKEL, UPHARSIN (מנא מנא תקל ופרסין) Words written by a mysterious hand on the walls of Belshazzar's palace, and interpreted by Daniel as predicting the doom of the King and his dynasty. The incident is described as follows: Once when King Belshazzar was banqueting with his lords and drinking wine from the golden vessels of the temple of YHWH, a man's hand was seen writing on the wall certain mysterious words. Frightened by the apparition, the King ordered his astrologers to explain the inscription; but they were unable to read it. Daniel was then summoned to the Royal Palace, and the King promised him costly presents if he could decipher the inscription. Daniel read it "Mene, mene, tekel, upharsin," and explained it to mean that God had "numbered" the Kingdom of Belshazzar and brought it to an end; that the King had been weighed and found wanting; and that his Kingdom was divided and given to the Medes and Persians.

The first question which presents itself to the critic, namely, why could the inscription be deciphered by Daniel only—engaged the attention of the Talmudists, who advanced various answers. Certain of them concluded that the Hebrew writing had been changed in the time of Ezra, so that even the Jews that were found in the royal court could not read an inscription written in archaic characters. But those who followed R. Simeon in maintaining that the writing had not changed found other solutions for the problem; *e. g.* it was written in the cryptographic combination את בש, each letter of each pair being substituted by its companion, *e. g.* יטת יטת אדך פרג חמט; or the words were written thus: ממתוס נעקפי אאלרן, one above the other, having to be read vertically; or אנם אנם לקת ניסרפו, each word backward; or again, נמא נמא קתל פורסין the first two letters of each word being transposed (Sanh. 22a). It is evident that the author of the Book of Daniel meant that the inscription was written in characters familiar to the King and wise men of Babylon, but that, as often happens with ancient inscriptions, the transposition of certain letters baffled every attempt to decipher them. . . ."

[2] See Lohr, Lenox R. and Friedman, William F., *Formulae for the solution of transposition ciphers*. Riverbank Publication No. 19, Geneva, Illinois, 1918.

[3] See Special Text No. 165, *Elementary Military Cryptography*, 1935, Sec. V. In this text the term "transposition rectangle" will be used to designate the matrix, frame, cage, or design regardless of whether the latter is completely filled or not.

c. In solving a cryptogram of this type the first step taken by the cryptanalyst is to ascertain the dimensions of the rectangle. Clues for this are usually afforded by finding the factors of the total number of letters in the cryptogram. Suppose the cryptogram contains 152 letters. The dimensions of the transposition rectangle may be 4×38 or 8×19, by which is meant that four hypotheses may be made with respect to its dimensions. The rectangle may consist of:

(1) 4 columns with 38 rows, *or*
(2) 38 columns with 4 rows, *or*
(3) 8 columns with 19 rows, *or*
(4) 19 columns with 8 rows.

In practical work it is rather unlikely to encounter a rectangle that conforms to hypothesis (1) or (2), and for the present these may be discarded. As to choosing between hypotheses (3) and (4), a rather simple test to be described presently will disclose which is the more probable.

d. It is obvious that if the cryptogram is transcribed within a rectangle of the correct dimensions, the letters in each row will be the ones which actually were in those rows in the original transposition rectangle and formed good plain text therein. *In fact, the rows of letters in the correctly-dimensioned rectangle would read plain text were it not for the transposition which they have undergone within the rows.* Therefore, the rows of a correctly-dimensioned rectangle are more likely to manifest the expected vowel-consonant proportions of normal plain text than are the rows of an incorrectly-dimensioned rectangle, because in the latter case there are brought into some of the rows letters which belong to other rows and which are likely to disturb the normal vowel-consonant proportions of plain text. That is, in an incorrectly-dimensioned rectangle some of the rows will have too many consonants and not enough vowels, in other rows this relationship will be reversed; whereas in a correctly-dimensioned rectangle each row will have the proper number of vowels and consonants. Hence in solving an unknown cryptogram of this type, if a count is made of the vowels and consonants in the rows of rectangles of various probable dimensions, that rectangle in which the rows show the best proportions of vowels and consonants is most likely to be correct, and the one that should be tried first.

e. Having ascertained the correct dimensions of the rectangle by the foregoing procedure, the next step is to experiment with the columns of the rectangle, trying to bring together several columns which will then show "good" digraphs, trigraphs, or polygraphs in the *rows* formed by juxtaposing the columns. This process of combining or matching columns in order to build up these fragments of plain text will herein be referred to as *anagramming.*[4]

f. The procedure is to select a column which has a good assortment of high-frequency letters and find another column which may be placed before or after the selected column to build up high-frequency digraphs in the rows; when such a pair of columns has been found, attempt is made to add another column before or after this pair to build up high-frequency trigraphs, and so on, gradually building up longer and longer polygraphs until entire words begin to appear in the respective rows of the rectangle. In this process of anagramming, advantage may be taken of simple mathematical considerations such as adding the normal plain-text frequency values of the digraphs in the rows to assist in discarding combinations which are on the borderline of choice. However, it must be noted that the totals obtained by simple addition of the frequency values of the digraphs should be considered only as rough approximations or guides in weighing probabilities in favor of one hypothesis as against another, for theoretically the probability of the simultaneous occurrence of two or more independent events is the *product*, and not the sum, of their respective probabilities. In most cases the calculation of products involves an amount of labor unwarranted by the results to be expected, so that simple addition of probabilities is usually sufficient. However, if tables of the logarithms of the probabilities are readily available, the addition of these logarithms becomes a simple matter and affords a more accurate guide in selection of combinations produced in the anagramming process.[5] Once a set of four or five columns has been correctly assembled it is usually the case that the process may be completed very quickly, for with the placement of each column the number of remaining columns possible for selection diminishes; toward the close of the process, when only two or three columns remain, their placement is almost automatic.

---

[4] The Standard Dictionary defines the word *anagram* as follows: "(noun) 1. The letters of a word or phrase so transposed as to make a different word or phrase; as, 'time' and 'mite' are *anagrams* of 'emit'. 2. A transposition; interchange." As a verb, it is defined as "to anagrammatize; to make an anagram of; make anagrams." (The construction of anagrams was a very widespread pastime in previous centuries. See Wheatley's *Of Anagrams*, London, 1862.) A strict interpretation of the word would therefore confine it to cases wherein the letters to be rearranged already form bonafide words or intelligible phrases. However, this would hardly be broad enough for cryptanalytic purposes. As used in cryptanalysis the word is commonly employed as a verb to refer to the process of rearranging the disordered letters of cipher text so as to reconstruct the original plain text.

g. It is desirable, as a final step, to try to reconstruct, if possible, the literal key from which the numerical transposition key was derived.

7. Example of solution.—*a.* Given the following cryptogram, the steps in solution will be set forth in detail:

CRYPTOGRAM (126 letters)

```
I L H H D   T I E O E   U D H T S   O N S O O   E E E E I   O E F T R
R H N E A   T N N V U   T L B F A   E D F O Y   C A P D T   R R I I A
R I V N L   R N R W E   T U T C U   V R A U O   O O F D A   O N A J I
U P O L R   S O M T N   F R A N F   M N D M A   S A F A T   Y E C F X
R T G E T   A
```

b. The cryptogram contains 126 letters (factors of 126: 2, 3, 6, 7, 9, 14, 18, 21), suggesting rectangles of 7×18 or 9×14. If the former dimensions are taken, the rectangle may have 7 columns and 18 rows or 18 columns and 7 rows; if the latter dimensions are taken, it may have 9 columns and 14 rows or 14 columns and 9 rows. The factors of 126 do not, of course, preclude the possibility that the rectangle may be 6×21, that is, with 21 columns and 6 rows or 6 columns and 21 rows. If no good results were obtained by testing rectangles of the dimensions indicated (7×18 or 9×14), then one would proceed to test rectangles 6×21. In the event that all tests on the basis of a completely-filled rectangle failed, then it would be assumed that the rectangle may be incompletely filled. In making the vowel-consonant test described in paragraph 6*d*, it is advisable to base the count on the columns as well as on the rows of a rectangle, since it is possible that the cryptogram was prepared by inscribing the plain text in rows and transcribing the text from the columns, or *vice versa*. After examining a rectangle both horizontally and vertically, it is often possible to discard various arrangements without further tests. For example, at A in figure 1 there is shown a rectangle of 7 columns and 18 rows. Now in a row of 7 letters there should be (7×40 percent=2.8) either 2 or 3 vowels; but rows 12 and 15 contain no vowels at all and rows 8 and 9 contain 5 vowels, row 16, 6 vowels. It is concluded at once that this arrangement is highly improbable. If the plain text had been inscribed vertically in this same rectangle, and then the rows had been transposed in forming the cipher text, then in each column (18 letters) there should be (18×40 percent=7.2) about 7 vowels; but column 2 contains 11 vowels and column 6 only 4. This likewise indicates that it is highly improbable that the message was inscribed vertically and the cryptogram formed by transposing the rows. But when the arrangement at

---

[5] A suggestion for which the author is indebted to Mr. A. W. Small, junior cryptanalyst in this office. The principle makes practicable the use of tabulating machinery for the purpose of speeding up and facilitating the matching of columns in the anagramming process.

B in figure 1 is studied, it is not so easy to say at once that it is improbable. For in 18 letters there should be about 7 vowels and none of the rows of this arrangement shows too great a departure from this expected number. This possibility will have to be explored further and it is for the moment put aside. If it be assumed that the message was inscribed vertically in the rectangle 18×7 and the rows subjected to transposition, there should be (7×40 percent=2.8) 2 or 3 vowels in each column. But since several of the columns show rather considerable departures from this expected number, it may be concluded that a vertical inscription and horizontal transcription is not probable and this assumption may be eliminated. Then the arrangements at C and D in figure 1 are studied in the same manner, with the result that at the end of the study the situation as regards the various assumptions is summarized as follows:

7 × 18

Row No.	1	2	3	4	5	6	7	Number of vowels
1	I	O	N	T	T	U	M	3
2	L	O	N	R	C	P	A	2
3	H	E	V	R	U	O	S	3
4	H	E	U	I	V	L	A	4
5	D	E	T	I	R	R	F	2
6	T	E	L	A	A	S	A	4
7	I	I	B	R	U	O	T	4
8	E	O	F	I	O	M	Y	5
9	O	E	A	V	O	T	E	5
10	E	F	E	N	O	N	C	3
11	U	T	D	L	F	F	F	1
12	D	R	F	R	D	R	X	0
13	H	R	O	N	A	A	R	3
14	T	H	Y	R	O	N	T	2
15	S	N	C	W	N	F	G	0
16	O	E	A	E	A	M	E	6
17	N	A	P	T	J	N	T	1
18	S	T	D	U	I	D	A	3
Number of vowels	7	11	6	6	10	4	7	

A

18 × 7

Row No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Number of vowels
1	I	E	S	E	T	T	B	Y	R	N	T	A	A	P	T	M	F	X	6
2	L	O	O	E	R	N	F	C	I	L	U	U	O	O	N	N	A	R	9
3	H	E	N	E	R	N	A	A	I	R	T	O	N	L	F	D	T	T	6
4	H	U	S	I	H	V	E	P	A	N	C	O	A	R	R	M	Y	G	7
5	D	D	O	O	N	U	D	D	R	R	U	O	J	S	A	A	E	E	9
6	T	H	O	E	E	T	F	T	I	W	V	F	I	O	N	S	C	T	6
7	I	T	E	F	A	L	O	R	V	E	R	D	U	M	F	A	F	A	8
Number of vowels	2	4	4	6	2	1	3	2	4	1	2	5	5	2	1	2	3	2	

B

9 × 14

Row No.	1	2	3	4	5	6	7	8	9	Number of vowels
1	I	S	T	B	R	T	A	T	F	2
2	L	O	R	F	I	U	O	N	A	5
3	H	N	R	A	I	T	N	F	T	2
4	H	S	H	E	A	C	A	R	Y	4
5	D	O	N	D	R	U	J	A	E	4
6	T	O	E	F	I	V	I	N	C	4
7	I	E	A	O	V	R	U	F	F	5
8	E	E	T	Y	N	A	P	M	X	4
9	O	E	N	C	L	U	O	N	R	4
10	E	E	N	A	R	O	L	D	T	4
11	U	I	V	P	N	O	R	M	G	3
12	D	O	U	D	R	O	S	A	E	5
13	H	E	T	T	W	F	O	S	T	2
14	T	F	L	R	E	D	M	A	A	3
Number of vowels	6	10	3	5	5	7	7	3	5	

C

14 × 9

Row No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Number of vowels
1	I	E	O	F	N	E	T	N	T	O	U	N	M	C	6
2	L	U	O	T	N	D	R	L	C	F	P	F	A	F	3
3	H	D	E	R	V	F	R	R	U	D	O	R	S	X	3
4	H	H	E	R	U	O	I	N	V	A	L	A	A	R	7
5	D	T	E	H	T	Y	I	R	R	O	R	N	F	T	4
6	T	S	E	N	L	C	A	W	A	N	S	F	A	G	4
7	I	O	I	E	B	A	R	E	U	A	O	M	T	E	10
8	E	N	O	A	F	P	I	T	O	J	M	N	Y	T	6
9	O	S	E	T	A	D	V	U	O	I	T	D	E	A	8
Number of vowels	4	3	9	2	2	4	4	2	5	5	3	1	5	2	

D

FIGURE 1.

Rectangle 7 × 18

7 columns and 18 rows:
(1) Horizontal inscription, columnar transcription........................ Very improbable.
(2) Vertical inscription, horizontal transcription........................ Very improbable.
18 columns and 7 rows:
(3) Horizontal inscription, columnar transcription........................ Possible.
(4) Vertical inscription, horizontal transcription........................ Improbable.

Rectangle 9 × 14

9 columns and 14 rows:
(5) Horizontal inscription, columnar transcription........................ Possible.
(6) Vertical inscription, horizontal transcription........................ Improbable.
14 columns and 9 rows:
(7) Horizontal inscription, columnar transcription........................ Improbable.
(8) Vertical inscription, horizontal transcription........................ Very improbable.

c. Discarding all assumptions except (3) and (5), the latter are subjected to further scrutiny. Suppose the average amount of deviation from the expected number of vowels in each row is calculated by finding the difference between the actual and expected numbers in each row, adding these differences (neglecting signs), and dividing by the total number of rows. For assumptions (3) and (5) the results are as follows:

18 × 7

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Number of vowels	Deviation from expected number
1	I	E	S	E	T	T	B	Y	R	N	T	A	A	P	T	M	F	X	6	1.2
2	L	O	O	E	R	N	F	C	I	L	U	U	O	O	N	N	A	R	9	1.8
3	H	E	N	E	R	N	A	A	I	R	T	O	N	L	F	D	T	T	6	1.2
4	H	U	S	I	H	V	E	P	A	N	C	O	A	R	R	M	Y	G	7	.2
5	D	D	O	O	N	U	D	D	R	R	U	O	J	S	A	A	E	E	9	1.8
6	T	H	O	E	E	T	F	T	I	W	V	F	I	O	N	S	C	T	6	1.2
7	I	T	E	F	A	L	O	R	V	E	R	D	U	M	F	A	F	A	8	.8

Total deviation = 8.2
Average deviation = 1.2

FIGURE 1e.

9×14

	1	2	3	4	5	6	7	8	9	Number of vowels	Deviation from expected number
1	I	S	T	B	R	T	A	T	F	2	1.6
2	L	O	R	F	I	U	O	N	A	5	1.4
3	H	N	R	A	I	T	N	F	T	2	1.6
4	H	S	H	E	A	C	A	R	Y	4	.4
5	D	O	N	D	R	U	J	A	E	4	.4
6	T	O	E	F	I	V	I	N	C	4	.4
7	I	E	A	O	V	R	U	F	F	5	1.4
8	E	E	T	Y	N	A	P	M	X	4	.4
9	O	E	N	C	L	U	O	N	R	4	.4
10	E	E	N	A	R	O	L	D	T	4	.4
11	U	I	V	P	N	O	R	M	G	3	.6
12	D	O	U	D	R	O	S	A	E	5	1.4
13	H	E	T	T	W	F	O	S	T	2	1.6
14	T	F	L	R	E	D	M	A	A	3	.6

Total deviation = 12.6
Average deviation = .9

FIGURE 1*f*.

The average amount of deviation for assumption (5) is only 0.9 as against 1.2 for assumption (3); therefore the former assumption is considered to be somewhat better than the latter and it will be tried first.

d. The columns of the rectangle shown in figure 1*f* are now to be cut apart and the procedure of anagramming applied. (For this it is best to have the cryptogram written on cross-section paper preferably with ½-inch squares for ease in handling.) Consider column 7, with the letter J in row 5; this letter, if it is a part of a word, must be followed by a vowel, which eliminates columns 1, 3, 4, and 5 as possibilities for placement on the right of column 7. Here are the digraphs formed by combining column 7 with columns 2, 6, 8, and 9, respectively, and the totals obtained by adding the frequency values of the digraphs formed in the rows:

(The frequencies shown are as given in table 6, appendix to *Military Cryptanalysis, Part I.*)

(1)			(2)			(3)			(4)		
7	2	Frequency value	7	6	Frequency value	7	8	Frequency value	7	9	Frequency value
A	S	41	A	T	47	A	T	47	A	F	4
O	O	6	O	U	37	O	N	77	O	A	7
N	N	8	N	T	82	N	F	9	N	T	82
A	S	41	A	C	14	A	R	44	A	Y	12
J	O	2	J	U	2	J	A	1	J	E	2
I	O	41	I	V	25	I	N	75	I	C	22
U	E	11	U	R	31	U	F	1	U	F	1
P	E	23	P	A	14	P	M	4	P	X	0
O	E	3	O	U	37	O	N	77	O	R	64
L	E	37	L	O	13	L	D	9	L	T	8
R	I	30	R	O	28	R	M	9	R	G	7
S	O	15	S	O	15	S	A	24	S	E	49
O	E	3	O	F	25	O	S	14	O	T	19
M	F	1	M	D	1	M	A	36	M	A	36
Total		262	Total		371	Total		427	Total		313

FIGURE 2.

Combination (3) gives the highest frequency value for the digraphs and an attempt is made to add a column to it. Here are some of the combinations tried:

7	8	1	7	8	2	7	8	3	7	8	9
A	T	I	A	T	S	A	T	T	A	T	F
O	N	L	O	N	O	O	N	R	O	N	A
N	F	H	N	F	N	N	F	R	N	F	T
A	R	H	A	R	S	A	R	H	A	R	Y
J	A	D	J	A	O	J	A	N	J	A	E
I	N	T	I	N	O	I	N	E	I	N	C
U	F	I	U	F	E	U	F	A	U	F	F
P	M	E	P	M	E	P	M	T	P	M	X
O	N	O	O	N	E	O	N	N	O	N	R
L	D	E	L	D	E	L	D	N	L	D	T
R	M	U	R	M	I	R	M	V	R	M	G
S	A	D	S	A	O	S	A	U	S	A	E
O	S	H	O	S	E	O	S	T	O	S	T
M	A	T	M	A	F	M	A	L	M	A	A

FIGURE 3.

e. Each of these combinations shows at least one "impossible" trigraph and several "poor" ones.[6] After more or less work along these lines, the cryptanalyst begins to get the feeling that "something is wrong," for, as a rule, once a correct start has been made in cases of this kind, solution comes rather quickly. Hence, the cryptanalyst decides here that possibly his first choice of combination (3) was a bad one, even though it gave the greatest total when frequency values for the digraphs were summed. The second greatest total was for combination (2) in which columns 7 and 6 were put together. The infrequent digraph J U suggests a word such as JUST or JUNCTION. If it were the former there should be a column containing an S in the 5th row, and there is no such column. If the word is JUNCTION, there should be a column containing an N in the 5th row, and there is only one such column, the 3d. Placing column 3 after columns 7–6 gives the trigraphs shown in figure 4–A. All of these trigraphs are excellent except the last, and that one may be either an abbreviation of a signature, or possibly nulls added to complete the rectangle. If the word JUNCTION is correct then there should be a column with a C in the 5th row; but none is found. However, column 9 has a C in the 6th row, and if it happened that the last column on the right is No. 3, then column 9 would be the 1st column. Thus, as shown in figure 4–B, the arrangement of columns becomes 9 . . . . . 7 6 3.

---

[6] Following the steps taken in subpar. *d*, frequency weights may be given the various trigraphs in fig. 3 and the sums obtained taken as indications of the relative probability of each of the four trials. These steps are here omitted, for they are obvious.

7	6	3
A	T	T
O	U	R
N	T	R
A	C	H
J	U	N
I	V	E
U	R	A
P	A	T
O	U	N
L	O	N
R	O	V
S	O	U
O	F	T
M	D	L

FIGURE 4–A.

9	?	?	?	?	?	7	6	3
F	.	.	.	.	.	A	T	T
A	.	.	.	.	.	O	U	R
T	.	.	.	.	.	N	T	R
Y	.	.	.	.	.	A	C	H
E	.	.	.	.	.	J	U	N
C	.	.	.	.	.	I	V	E
F	.	.	.	.	.	U	R	A
X	.	.	.	.	.	P	A	T
R	.	.	.	.	.	O	U	N
T	.	.	.	.	.	L	O	N
G	.	.	.	.	.	R	O	V
E	.	.	.	.	.	S	O	U
T	.	.	.	.	.	O	F	T
A	.	.	.	.	.	M	D	L

FIGURE 4–B.

9	1	5	2	8	4	7	6	3
F	I	R	S	T	B	A	T	T
A	L	I	O	N	F	O	U	R
T	H	I	N	F	A	N	T	R
Y	H	A	S	R	E	A	C	H
E	D	R	O	A	D	J	U	N
C	T	I	O	N	F	I	V	E
F	I	V	E	F	O	U	R	A
X	E	N	E	M	Y	P	A	T
R	O	L	E	N	C	O	U	N
T	E	R	E	D	A	L	O	N
G	U	N	I	M	P	R	O	V
E	D	R	O	A	D	S	O	U
T	H	W	E	S	T	O	F	T
A	T	E	F	A	R	M	D	L

FIGURE 5.

f. It is believed that the procedure has been set forth with sufficient detail so as to make further demonstration unnecessary. The rectangle can be completed very quickly and is found to be as shown in figure 5.

g. It will be interesting to see if a calculation based upon the sum of the logarithms of the probabilities given in figure 2 would have given the correct combination as the first choice. Note the results shown in figure 6. This calculation gives the correct combination as first choice, *viz*, 7–6, with a logarithmically-weighted value of 17.35 as against a value of 16.51 for combination 7-8, which was the first one tried on the basis of merely the sums of the frequency values of the digraphs.

(1)			(2)			(3)			(4)		
7 2	Frequency value	Logarithms	7 6	Frequency value	Logarithms	7 8	Frequency value	Logarithms	7 9	Frequency value	Logarithms
A S	41	1.61	A T	47	1.67	A T	47	1.67	A F	4	0.60
O O	6	.78	O U	37	1.57	O N	77	1.89	O A	7	.85
N N	8	.90	N T	82	1.91	N F	9	.95	N T	82	1.91
A S	41	1.61	A C	14	1.15	A R	44	1.64	A Y	12	1.08
J O	2	.30	J U	2	.30	J A	1	.00	J E	2	.30
I O	41	1.61	I V	25	1.40	I N	75	1.88	I C	22	1.34
U E	11	1.04	U R	31	1.49	U F	1	.00	U F	1	.00
P E	23	1.36	P A	14	1.15	P M	4	.60	P X	0	−1.00
O E	3	.48	O U	37	1.57	O N	77	1.89	O R	64	1.81
L E	37	1.57	L O	13	1.11	L D	9	.95	L T	8	.90
R I	30	.48	R O	28	1.45	R M	9	.95	R G	7	.85
S O	15	1.18	S O	15	1.18	S A	24	1.38	S E	49	1.69
O E	3	.48	O F	25	1.40	O S	14	1.15	O T	19	1.28
M F	1	.00	M D	1	.00	M A	36	1.56	M A	36	1.56
Total	262	13.40	Total	371	17.35	Total	427	16.51	Total	313	13.17

FIGURE 6.

As a matter of interest, it may be observed that the combination 7–6 is 7 times more probable than combination 7–8, since the difference between 17.35 and 16.51 is .84, which is the logarithm of 7. Likewise, combination 7–6 is roughly 15,000 times more probable than combination 7–9, since 17.35−13.17=4.18.

8. The probable-word method of solution.—*a.* The probable-word method of attack is as important in the solution of transposition ciphers as it is in the solution of substitution ciphers, and if the cryptanalyst is able to assume the presence of such probable words as are usually encountered in military communications, the solution, as a rule, comes very quickly.

b. As an illustration, looking at the first row of letters in the rectangle shown in figure 1*f*, the letters I S T B R T A T F almost at once suggest FIRST BATTALION as the initial words of the message. A rearrangement of the columns of the cryptogram to bring the necessary letters into juxtaposition at once discloses the key. Thus:

```
9 1 5 2 8 4 7 6 3
F I R S T B A T T
A L I O N
```

It will be noted that this assumption requires that there be a column headed by F A, another headed by I L, another headed by R I, and so on. Had such columns not been found, then the word BATTALION would not be possible. In that case the word FIRST would still remain as a point of departure for further experimentation.

c. In the foregoing illustration, the probable word was assumed to appear in the first line of text in the rectangle. If the probable word being sought is in the interior of the message, the steps must be modified somewhat but the basic principle remains unchanged. The modifications are of course obvious.

9. General remarks on solution.—*a.* In solving transposition ciphers advantage should be taken of all the characteristics and idiosyncrasies which are applicable to the language of the enemy, because they often afford clues of considerable assistance to the cryptanalyst. In all languages there are certain letters, usually of medium or low frequency, which combine with other letters to form digraphs of high frequency. For instance, in English the letter H is of medium frequency, but it combines with T to form the digraph T H, which is of highest frequency in literary text; it also combines with C, a letter of medium frequency, to form the fairly frequent digraph C H. The letter V is almost in the low-frequency category yet it combines with E to form the digraph V E, which in military text is the 14th in frequency. The low-frequency letter K often combines with C to form the digraph C K. Consequently, in working with transposition ciphers in English, when there is an H, attempts should be made to combine it first with a T or with a C; a V should be combined first with an E; a K should be combined first with a C; and so on.

b. There is usually in every language at least one letter which can be followed by only a certain other letter, forming what may be termed an *obligatory sequence*, or an *invariable digraph.* In all languages having the letter Q, the combination Q U constitutes such an invariable digraph.[7] In bonafide words of the German language the letter C is never used by itself; when present the letter C is invariably followed by an H, except on rare occasions when the digraph C K is employed. In English, the letter J can be followed only by a vowel; the letter X can only be preceded by a vowel and, except at the end of a word, can only be succeeded by a vowel, or by one of a limited number of consonants (C H P T), and so on. Letters which behave in this manner, that is, letters which have what may be called a *limited affinity* in combining with other letters to form digraphs, constitute good points of departure for solution and are therefore of sufficient importance to warrant their being designated by the more or less descriptive name of *pilot letters.*

c. The presence of pilot letters in a transposition cipher often forms the basis for the assumption of probable words. Obviously, a special lookout should be kept for *words* of rather high frequency (in military correspondence) which contain *letters* of low or medium frequency. The frequent word CAVALRY, for example, would suggest itself if the cryptogram has the letters C, V, L, and Y, which are all of medium frequency. The important word ATTACK suggests itself if the cryptogram has a K, a letter of low frequency, and a C, one of medium frequency; and so on.

d. The mechanics of simple columnar transposition make possible the production of rather long sequences of vowels and long sequences of consonants in the text of the cryptogram. Note, for example, in the cryptogram on p. 6, the sequence of vowels O O E E E E I O E, and the sequence of consonants V N L R N R W. If the enciphering or plain-text rectangle is consulted, it will be seen that these two sequences belong together, that is, they are in adjacent columns in that rectangle. It is a characteristic of plain text that consonant-vowel or vowel-consonant digraphs are much more frequent than consonant-consonant or vowel-vowel digraphs,[8] and therefore when long sequences of consonants and of vowels are found in transposition ciphers, a good start toward solution may result from assuming that such sequences come from adjacent columns.

e. It should, however, be noted in connection with tell-tale letters such as Q (entering into the composition of Q U) and C (entering into the composition of C H), that astute cryptographers who realize the clues which such letters afford often replace invariable digraphs by single characters, usually those rarely used in the language in question. For example, C H in German may be replaced by Q, Q U in French, by K, and so on. When this is done, solution is made more difficult; but only in those cases where it is dependent upon finding letters forming obligatory sequences in plain text does this sort of subterfuge become a factor of importance.

---

[7] The letter Q may, of course, be part of an abbreviation, such as SQ for "square," or it may be used as a null, or as a sign of punctuation. However, unless there are good reasons for believing that this letter is used for such purposes, QU may be considered to be an invariable digraph.

[8] The CV and VC digraphs constitute about 62 percent of all digraphs.

f. The presence of many Q's, or K's, or X's in a transposition cipher should not, however, be taken as *prima facie* evidence of the type of replacement noted in the preceding subparagraph. It is possible that such letters may be used as sentence separators or other punctuation; possibly they may be nulls, although the alert cryptographer would either use nulls not at all or, if he had to, would use letters of medium or high frequency for this purpose.

g. Because it is important that the cryptanalyst take advantage of every peculiarity specifically applicable to a cryptogram to be solved, especially as regards the presence of low-frequency letters, it is advisable that a uniliteral frequency distribution be prepared, just as though he were going to deal with a substitution cipher. This is probably the quickest way of bringing to light the peculiarities which may be helpful in solution.

10. Reconstruction of literal key.—*a.* The reconstruction or recovery of the literal key from which the numerical transposition key was derived is naturally the last step in the solution of cryptograms of this type. It is often of more than merely academic interest, because if it is found that the enemy is employing for this purpose some well-known book, or words or phrases of a simple nature associated with the locale of operations, this fact may be of highest importance in subsequent work.

b. In this process there are only a few guiding principles to be noted and much must be left to the ingenuity and imaginative powers of the cryptanalyst. Taking as an example the numerical key uncovered in the solution of the cryptogram in paragraph 7, the procedure will be set forth below.

c. The numerical key referred to was found to be 9 1 5 2 8 4 7 6 3. Assuming that this sequence was derived in the usual manner, by assigning numbers to the letters of a key word in accordance with their relative positions in the normal alphabet, the sequence forms the basis for the *key-word reconstruction diagram* shown in figure 7–A, in which the individual key numbers are written from left to right on different "levels" so that each level contains only numbers normally in succession.

	9	1	5	2	8	4	7	6	3
1		1		2					3
2						4			
3			5					6	
4							7		
5					8				
6	9								

FIGURE 7–A.

	9	1	5	2	8	4	7	6	3
1		ABCDE 1		ABCDE 2					ABCDE 3
2						FGHIJ 4			
3			KLMNO 5					KLMNO 6	
4							LMNOP 7		
5					MNOPR 8				
6	R-Z 9								

FIGURE 7–B.

d. It is likely that the digit 1 on the first level in the key-word reconstruction diagram represents a letter at or at least close to the beginning of the alphabet. Since the digits 2 and 3 are on the same level as the digit 1, it is likely (1) that the letter represented by 1 occurs 2 more times in the key word, or (2) that the digit 2 represents another letter, also near the beginning of the alphabet, and that this letter is repeated, or (3) that the digits 2 and 3 represent 2 different letters both near the beginning of the alphabet, or (4) that all three digits represent different letters but all near the beginning of the alphabet. The digit 4, on the second level in the reconstruction diagram, must represent a letter beyond the letter represented by the digit 3; the digit 5 must represent one beyond the letter represented by the digit 4, and the digit 6 may represent the same letter as 5, or a letter not much beyond that represented by 5. Assuming that the letters composing the key word are fairly well distributed over the entire alphabet, the digit 7 must represent a letter near or slightly beyond the middle of the alphabet, the digit 8 must represent one further toward the end of the alphabet than does the digit 7, and so on. Assigning several values to the digits, in accordance with the foregoing principle, the results are as shown in figure 7–B.

e. It is perhaps possible that some students may find the process of reconstructing the literal key somewhat easier if the variant possible letters are merely listed under the respective key numbers as shown in figure 7–C. The candidates for the successive positions in the literal key thus appear in a rather condensed space and the eye is able to pick up "good" combinations very quickly.

9	1	5	2	8	4	7	6	3
R	A	K	A	M	F	L	K	A
S	B	L	B	N	G	M	L	B
T	C	M	C	O	H	N	M	C
U	D	N	D	P	I	O	N	D
V	E	O	E	R	J	P	O	E
W								
X								
Y								
Z								

FIGURE 7-C.

f. Now comes the trying process of finding a "good" word in this assemblage of letters. The beginning and end of the word are the easiest points of attack, and it is useful to keep in mind the relative frequency order of letters as initial and final letters of the language in question. For English, the data are as follows: [9]

As initial letters..............T S A F C O R D N P E M I W B H L U G Y V J Q K Z X
As final letters.................E T D S N Y R O H L A F G P M X C K W U B I Z Q J V

Studying the candidate letters at the end of the key, it is seen that E is one of the possibilities. If that is correct, then a good ending would be one of the type vowel-consonant-vowel, with E as the final letter. There is but 1 vowel in the fourth level in the column under the digit 7, the letter O. This gives O K E, O L E, O M E, O N E as possible terminal trigraphs, the best of which from a frequency standpoint is ONE. Seeing the letters P and H in columns 8–4, the ending PHONE and then the word TELEPHONE suggests itself. Checking to see if there are any inconsistencies, none is found and the solution is:

Numerical key......9 1 5 2 8 4 7 6 3
Literal key...........T E L E P H O N E

---

[9] Taken from Tables 2–D (2) and 2–E (2), p. 111, *Military Cryptanalysis, Part I.*

g. In future studies, cases will be encountered wherein the reconstruction of the numerical key is an essential or, at least, a useful element after the solution of one or more cryptograms has been achieved by cryptanalysis. This is done in order that subsequent cryptograms in the same key can be read directly without cryptanalysis. The reconstruction of the numerical key is, however, a different process than the one illustrated in this paragraph, wherein the problem is solely one of building up a literal key from its numerical equivalent. The purpose in reconstructing the literal key is to give clues as to the *source* from which keys are derived or taken. Sometimes this may lead to ascertaining a book which is used for this purpose and which may be available by purchase at bookshops; or it may be a well-known document, a telephone directory, etc. Obviously, if the source document or book can be located the solution of future cryptograms in the same system becomes merely a matter of decipherment and such cryptograms no longer form the material for cryptanalytic efforts. The method of reconstructing the literal key is, obviously, easier to apply in the case of long numerical keys than in the case of short ones; in general, the longer the numerical key the easier is the recovery of the literal key.

11. Column and row transposition.—It should be obvious that when the rows as well as the columns of a completely-filled rectangle undergo transposition the increase in security is hardly worth mention, since the underlying procedure in solution aims simply at assembling a few columns on the basis of "good" digraphs and trigraphs brought to light by juxtaposing *columns*. After three or four columns have been properly juxtaposed, the placement of additional columns becomes easier and easier, merely by continuing to build upon the fragments of words *in the rows*. Hence, the cryptanalyst is, during a large part of the process, not particularly interested in the intelligibility of the text he is building up; only at the end of the process does this become a factor. When all of the columns have been assembled in proper order, then the text will read continuously in the normal manner (left to right, top to bottom). If it does not, then it is usually a very simple matter to rearrange the rows of the rectangle to bring this about, since the letters at the ends and beginnings of the rows give the necessary clues for continuity.

SECTION III

INCOMPLETELY—FILLED RECTANGLES

Paragraph

General principles underlying solution 12
Delimiting the lengths of the columns of the rectangle; constructing the "hat" diagram 13
Solution of example 14
Alternative method of solution 15
Concluding remarks on simple columnar transposition 16

12. General principles underlying solution.—*a.* In the system designated keyed columnar transposition the feature which differentiates an incompletely-filled rectangle from one that is completely filled is a very simple one from the cryptographic point of view: The bottom row of the rectangle in the former case merely lacks one or more letters, a feature which only very slightly complicates the system in practical operation. But the consequences of this simple difference between the two types are, from the cryptanalytic point of view, quite profound, and the cryptanalytic effect of this small change in cryptographic procedure is seemingly all out of proportion with the simplicity of the difference.

b. Cryptograms involving completely-filled rectangles are rather easy to solve because of two circumstances. In the first place, since the rectangle is completely filled, the various possible dimensions of the rectangle can be ascertained by noting the factors of the total number of letters. Usually only a few possibilities are indicated and therefore this materially reduces the amount of experimentation that would be required in the absence of this situation, since it is obvious that when working with incompletely-filled rectangles a good many rectangles of various dimensions become possibilities for trial. In the second place, the columns in a completely-filled rectangle all contain the same number of letters, and therefore the anagramming process (matching and assembling of columns) can be performed without any mental reservations such as must be made in working with incompletely-filled rectangles because of uncertainty as to whether the letters which are juxtaposed to form digraphs and trigraphs really come from the same row in the plain-text rectangle. The latter statement calls for a bit more explanation.

c. The *columns* of an incompletely-filled rectangle are of two sorts which may conveniently be designated as *long* and *short.* The long columns are at the left of the rectangle and each one contains just one more letter than the short columns, which are at the right. This follows, of course, from the fact that it is only the last row in such a rectangle which lacks one or more letters to complete the rectangle. The term *width,* as applied to a transposition rectangle, will be convenient to designate the number of columns, which is, of course, determined by the length of the numerical key or the number of letters in the literal key. Given the width of the rectangle and the total number of letters in the cryptogram, the length and number of the long and the short columns may be found by a simple calculation: Multiply the width of the rectangle by the smallest number which will yield a product greater than the total number of letters in the cryptogram. The multiplier gives the length of the long columns; this multiplier minus 1 gives the length of the short columns; the excess over the total number of letters gives the number of short columns; the latter deducted from the width gives the number of long columns. Thus, with a cryptogram of 287 letters and a rectangle 15 columns in width: $[(15\times20)-13=287]$ the long columns will have 20 letters, the short ones, 19 letters; there will be 13 short columns and 2 long ones.

d. Now if the cryptanalyst were able to cut up the text of a cryptogram produced from an incompletely-filled rectangle into sections corresponding in length with the actual long and short columns, he could handle these columns in exactly the same manner that he handles the equal-length columns in the solution of a cryptogram produced from a completely-filled rectangle. In fact, the solution would be easier because he knows that all the short columns fall at the right, all the long columns at the left of the transposition rectangle, and therefore the amount of experimentation he must undertake in his attempts to juxtapose columns in the anagramming process is considerably reduced. But, unfortunately, there is usually no way in which, at the initial stage of solution, the cryptanalyst can find out, from a single cryptogram, which are the long columns and which the short. This is obviously a matter directly connected with the specific transposition key, and the latter is the sole unknown factor in the whole problem.

e. If it were practicable to transcribe a cryptogram of this type according to *all* the possible transposition keys for a given width of rectangle, solution would obviously merely consist in scanning the various rectangles to find the one which is correct—for there will be only one such rectangle. A rectangle 15 columns in width may have been enciphered by any one of factorial 15 transposition keys.[1] While it is conceivable that machinery might be devised for this purpose, so that the production of the millions of possible rectangles could be effected in a relatively short time, in the present state of the art no such machinery has yet been devised. Furthermore, it is problematical whether a solution by such means could be achieved in a reasonable length of time even if the machinery were available, because of the immensity of the task it would have to perform.[2]

f. However, this question may be asked: Given a cryptogram of t letters and a rectangle of n columns in width, is it possible to transcribe the text within a single rectangle so that the latter will show what letters will constitute the respective columns for all possible transposition keys of n elements? If so, then such a rectangle would be useful in trying to solve the cryptogram, because the rectangle would then limit the amount of experimentation that would have to be performed by the anagramming process, since it would show whether or not two letters which are brought together in that process to form a digraph could possibly have been in the same row in the plain-text rectangle. If not, then of course there would be no use in forming such digraphs, and thus the number of trials becomes much reduced. Another way of indicating what is meant is to say that such a rectangle would show the maximum amount that one column may be shifted up or down in trying to match it with another column in the anagramming process. This will be made clearer in a subsequent paragraph. At this point it will merely be stated that it is easy to prepare a rectangle of the nature indicated above for any keyed, columnar-transposition cryptogram.

---

[1] Factorial 15, or $15 \times 14 \times 13 \times \ldots \times 1$, equals 1,369,944,576,000 different transposition keys.

[2] It is nevertheless pertinent to indicate that machinery for facilitating the "matching" or anagramming of columns has been devised and found to be quite practical in the solution of problems involving columnar transposition.

13. Delimiting the lengths of the columns of the rectangle; constructing the "hat" diagram.—*a*. Given the following cryptogram of 224 letters and an assumed width of 12 columns in the enciphering rectangle:

CRYPTOGRAM

ODNNP TIRNT DTURO EXALN IETGN WTTME
DSTEO ITDMA NLNOE BOUHE NLESE AACTR
MSCLC SOEFC FFTEE EMIAI TEAIJ NSOIV
FMBIE HBVTB ESRSY LXROR UMETY OIKNK
TNDAH IRHQI ETETN OTRAA VRIRS TGSEF
EAOOT HEACN SHEEV TRESR AIIEA TEEAL
AENEE MYTFI TANLN NUACL RENRT RATSO
ALODI RORYN NRGY

DISTRIBUTION

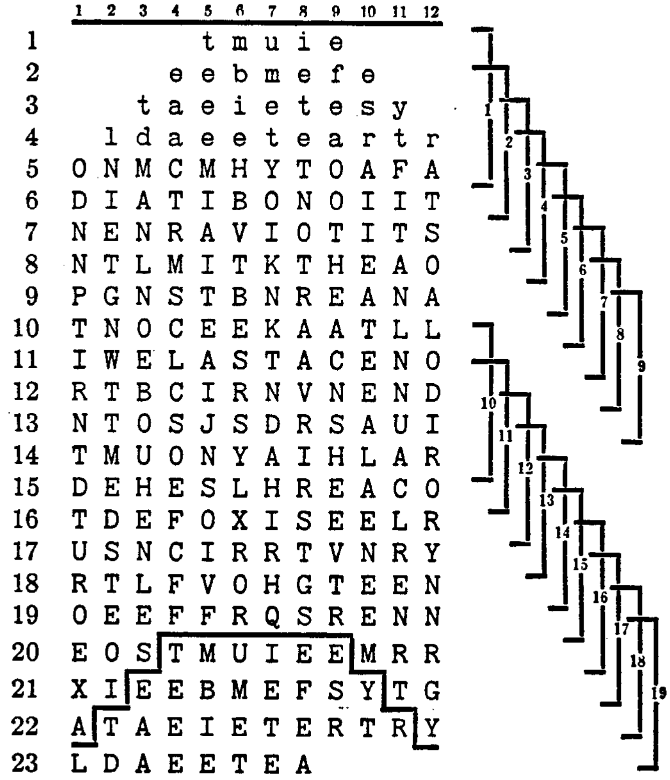
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
19	4	6	6	31	6	3	6	16	1	2	9	7	20	15	1	1	18	12	25	4	4	1	2	5	0	(224)

b. A cryptogram of 224 letters and a rectangle of 12 columns [(12×19)−4=224] indicates 4 short columns of 18 letters and 8 long columns of 19 letters. The outlines of a rectangle of this specification are drawn on a sheet of cross-section paper and the text is transcribed within it, for the moment assuming only that the transposition key consists merely of the straight sequence of numbers 1 to 12. Thus:

1	2	3	4	5	6	7	8	9	10	11	12
O	N	M	C	M	H	Y	T	O	A	F	A
D	I	A	T	I	B	O	N	O	I	I	T
N	E	N	R	A	V	I	O	T	I	T	S
N	T	L	M	I	T	K	T	H	E	A	O
P	G	N	S	T	B	N	R	E	A	N	A
T	N	O	C	E	E	K	A	A	T	L	L
I	W	E	L	A	S	T	A	C	E	N	O
R	T	B	C	I	R	N	V	N	E	N	D
N	T	O	S	J	S	D	R	S	A	U	I
T	M	U	O	N	Y	A	I	H	L	A	R
D	E	H	E	S	L	H	R	E	A	C	O
T	D	E	F	O	X	I	S	E	E	L	R
U	S	N	C	I	R	R	T	V	N	R	Y
R	T	L	F	V	O	H	G	T	E	E	N
O	E	E	F	F	R	Q	S	R	E	N	N
E	O	S	T	M	U	I	E	E	M	R	R
X	I	E	E	B	M	E	F	S	Y	T	G
A	T	A	E	I	E	T	E	R	T	R	Y
L	D	A	E	E	T	E	A				

FIGURE 8.

c. The rectangle shown in figure 8 is the same as though it had been assumed that the key numbers 9, 10, 11, and 12 happened to fall at the extreme right in the numerical transposition key. Columns 1 to 8, inclusive, would then be long columns, and columns 9, 10, 11, and 12 would be short columns. But suppose that the key numbers on the extreme right happened to be 1, 2, 3, and 4, instead of 9, 10, 11, and 12. Then columns 1, 2, 3, and 4 would be the short columns, 5 to 12 the long ones. In this case, making reference to figure 8, the final letter of column 1 would pass to the top of column 2; the final 2 letters of column 2 would pass to the top of column 3; the final 3 letters of column 3 would pass to the top of column 4; the final 4 letters of columns 4, 5, 6, 7, and 8 would pass to the top of columns 5, 6, 7, 8, and 9; the final 3 letters of column 9 would pass to the top of column 10; the final 2 letters of column 10 would pass to the top of column 11; and the final letter of column 11 would pass to the top of column 12. The results of the foregoing reasoning are embodied in the matrix or diagram shown in figure 9.



	1	2	3	4	5	6	7	8	9	10	11	12
1					t	m	u	i	e			
2				e	e	b	m	e	f	e		
3			t	a	e	i	e	t	e	s	y	
4		l	d	a	e	e	t	e	a	r	t	r
5	O	N	M	C	M	H	Y	T	O	A	F	A
6	D	I	A	T	I	B	O	N	O	I	I	T
7	N	E	N	R	A	V	I	O	T	I	T	S
8	N	T	L	M	I	T	K	T	H	E	A	O
9	P	G	N	S	T	B	N	R	E	A	N	A
10	T	N	O	C	E	E	K	A	A	T	L	L
11	I	W	E	L	A	S	T	A	C	E	N	O
12	R	T	B	C	I	R	N	V	N	E	N	D
13	N	T	O	S	J	S	D	R	S	A	U	I
14	T	M	U	O	N	Y	A	I	H	L	A	R
15	D	E	H	E	S	L	H	R	E	A	C	O
16	T	D	E	F	O	X	I	S	E	E	L	R
17	U	S	N	C	I	R	R	T	V	N	R	Y
18	R	T	L	F	V	O	H	G	T	E	E	N
19	O	E	E	F	F	R	Q	S	R	E	N	N
20	E	O	S	T	M	U	I	E	E	M	R	R
21	X	I	E	E	B	M	E	F	S	Y	T	G
22	A	T	A	E	I	E	T	E	R	T	R	Y
23	L	D	A	E	E	T	E	A				

FIGURE 9.

d. Now the capital letters in this matrix or diagram, which is often called a *crown* or *hat* diagram,[3] figure 9, represent the latters which are in the columns in case the first hypothesis (key numbers 9, 10, 11, 12 at the extreme right) is true. The capital letters above the heavy black line together with the lower-case letters at the top of the diagram represent the letters which are in the columns in case the second hypothesis (key numbers 1, 2, 3, 4 at the extreme right) is true. Therefore, since the hat diagram covers the two possible extremes with reference to the positions occupied by the short columns and embraces all possible intermediate conditions by showing what letters may be in the respective columns under *any* possible arrangement of long and short columns, *the hat diagram is applicable to any possible numerical key for the cryptogram in question and for the assumed width of rectangle.* Therefore, in the anagramming process the hat diagram shows the maximum possible amount that any column may be shifted up or down in juxtaposing two columns to form digraphs of letters assumed to come from the same row in the plain-text rectangle. This is because all the letters of the first row of the actual enciphering rectangle will be found in rows 1 to 5, inclusive, of figure 9; all the letters of the second row of the rectangle will be found in rows 2 to 6, inclusive, and so on, as indicated by the braces at the right in figure 9.

---

[3] Because the lower-case letters at the top form what is usually called the *crown* or *hat*.

e. Thus there arises the following important principle: Designating the number of short columns in a specific diagram by n, only such letters as fall within $(n+1)$ consecutive rows, will be letters that may have appeared in the same row in the actual transposition rectangle. Or, another way of stating the principle is this: Both members of any pair of letters actually in the same row in the transposition rectangle will be found only among the letters appearing in $(n+1)$ consecutive rows in the correct hat diagram. In the case under discussion, if the first letter of such a pair is located in row 8, for example, the other letter cannot be in rows 1, 2, 3, or 13 to 23 of figure 9.

f. The usefulness of this principle in connection with the construction and employment of the hat diagram will soon become apparent. For example, again referring to figure 9, take the letter Q in row 19, column 7; it must be followed by a U in the plain text. There are 4 U's in the message; they are in row 13 column 11, row 14 column 3, row 17 column 1, and row 20 column 6. Now the question is, can any of these 4 U's follow the U, or may one or more of them be eliminated from consideration at once? Since the U's in rows 13 and 14 fall outside the 4 consecutive rows above that in which the U is located, it follows that neither one of these U's can be the one that succeeds the U. Thus two candidates are automatically eliminated from consideration. The U in row 17 and the U in row 20 are both possible candidates.

14. Solution of example.—*a.* With the foregoing preliminaries out of the way, the solution of the cryptogram can now be carried forward with rapid progress. It has been indicated that the Q in row 19, column 7 (fig. 9), may be combined with either the U in row 17 column 1, or the U in row 20 column 6. Suppose the columns of figure 9 are now cut apart for ease in anagramming. Juxtaposing the indicated columns yields what is shown in figure 10. Since the combination shown at *a* in figure 10 involves column 1, it obviously begins with the letter O and ends with the letter A or L; no other letters can be added to this column. Since column 7 is already the maximum length this column can be under any circumstances, no letters can be added to it at the bottom. Therefore, all the digraphs possible to form by juxtaposing these two columns are indicated in figure 10*a*. There are only 17 digraphs in all, whereas there should be at least 18.

7	1
u	
m	
e	
t	
Y	
O	
I	O
K	D
N	N
K	N
T	P
N	T
D	I
A	R
H	N
I	T
R	D
H	T
Q	U
I	R
E	O
T	E
E	X
	A
	L

a

7	6
u	b
m	i
e	e
t	H
Y	B
O	V
I	T
K	B
N	E
K	S
T	R
N	S
D	Y
A	L
H	X
I	R
R	O
H	R
Q	U
I	M
E	E
T	T
E	

b

4	7	6
e		
a	u	b
a	m	i
c	e	e
T	t	H
R	Y	B
M	O	V
S	I	T
C	K	B
L	N	E
C	K	S
S	T	R
O	N	S
E	D	Y
F	A	L
C	H	X
F	I	R
F	R	O
T	H	R
E	Q	U
E	I	M
E	E	E
	T	T
	E	

c

1	12	4	7	6	10
		a	u	b	
		a	m	i	e
O	r	c	e	e	s
D	A	T	t	H	r
N	T	R	Y	B	A
N	S	M	O	V	I
P	O	S	I	T	I
T	A	C	K	B	E
I	L	L	N	E	A
R	O	C	K	S	T
N	D	S	T	R	E
T	I	O	N	S	E
D	R	E	D	Y	A
T	O	F	A	L	L
U	R	C	H	X	A
R	Y	F	I	R	E
O	N	F	R	O	N
E	N	T	H	R	E
X	R	E	Q	U	E
A	G	E	I	M	M
L	Y	E	E	E	Y
			T	T	T
			E		

d

FIGURE 10.

Hence, combination 7–1 is impossible, and combination 7–6 is the only one that needs to be considered further. There are many excellent digraphs in it, and only one which admittedly looks rather bad, the H X. Seeing the digraphs K B and K S in these columns, a good assumption to make is that the K's are preceded by the letter C. Is there a column with 2 C's in approximately the correct region? Column 4 meets this requirement. Note the excellent trigraphs it yields, as shown in figure 10*c*. It now becomes fairly easy to add columns to this nucleus. For instance, the trigraph R Y B suggests a word ending in R Y, such as INFANTRY, ARTILLERY, CAVALRY; the trigraph M O V suggest MOVING; the trigraph C K B suggests the word ATTACK; followed by a word beginning with B, and so on. Trial of only a few columns soon yields what is shown in figure 10*d*, from which it soon becomes probable that the long columns end with column 12, since the letters after L Y yield an impossible sequence (E E E Y). Since it was originally assumed that there are only 4 short columns in the transposition rectangle, and since 4 columns have already been placed at the right (4–7–6–10), the rectangle, with the columns thus far placed, must be as shown in figure 10*e*. This, then, at once tells what the limits of columns 2, 3, 5, 8, 9, and 11 must be, and the rectangle can now be filled in without further delay. The completed rectangle is shown in figure 11.

							1	12	4	7	6	10
1							O	R	C	E	E	S
2							D	A	T	T	H	R
3							N	T	R	Y	B	A
4							N	S	M	O	V	I
5							P	O	S	I	T	I
6							T	A	C	K	B	E
7							I	L	L	N	E	A
8							R	O	C	K	S	T
9							N	D	S	T	R	E
10							T	I	O	N	S	E
11							D	R	E	D	Y	A
12							T	O	F	A	L	L
13							U	R	C	H	X	A
14							R	Y	F	I	R	E
15							O	N	F	R	O	N
16							E	N	T	H	R	E
17							X	R	E	Q	U	E
18							A	G	E	I	M	M
19							L	Y				

FIGURE 10e.

	8	2	5	3	11	9	1	12	4	7	6	10
1	E	N	E	M	Y	F	O	R	C	E	E	S
2	T	I	M	A	T	E	D	A	T	T	H	R
3	E	E	I	N	F	A	N	T	R	Y	B	A
4	T	T	A	L	I	O	N	S	M	O	V	I
5	N	G	I	N	T	O	P	O	S	I	T	I
6	O	N	T	O	A	T	T	A	C	K	B	E
7	T	W	E	E	N	H	I	L	L	N	E	A
8	R	T	A	B	L	E	R	O	C	K	S	T
9	A	T	I	O	N	A	N	D	S	T	R	E
10	A	M	J	U	N	C	T	I	O	N	S	E
11	V	E	N	H	U	N	D	R	E	D	Y	A
12	R	D	S	E	A	S	T	O	F	A	L	L
13	I	S	O	N	C	H	U	R	C	H	X	A
14	R	T	I	L	L	E	R	Y	F	I	R	E
15	S	E	V	E	R	E	O	N	F	R	O	N
16	T	O	F	S	E	V	E	N	T	H	R	E
17	G	I	M	E	N	T	X	R	E	Q	U	E
18	S	T	B	A	R	R	A	G	E	I	M	M
19	E	D	I	A	T	E	L	Y				

FIGURE 11.

b. The last step, recovering the literal key, is then taken. The key is to be found among the letters of the literal key reconstruction diagram in figure 12.

8	2	5	3	11	9	1	12	4	7	6	10
						ABC 1					
	DEF GHI 2		DEF GHI 3					DEF GHI 4			
		JKL MN 5								JKL MN 6	
									MNO PQ 7		
NOP RST 8					NOP RST 9						NOP RST 10
				R-Z 11			R-Z 12				

FIGURE 12.

The termination ATIONS seems a likely possibility. If this is correct, assignment of letters becomes modified as shown in figure 13:

8	2	5	3	11	9	1	12	4	7	6	10
						A 1					
	DEF GHI 2		DEF GHI 3					I 4			
		JKL MN 5								N 6	
									O 7		
PRS 8					PRS 9						S 10
				T 11			T 12				

FIGURE 13.

The word PENETRATIONS will fit and it is taken to be presumably correct. There is no absolute certainty about the matter, for it is conceivable and possible that there are other words which can be made to fit the sequence of key numbers given.

15. Alternative method of solution.—*a.* The foregoing solution will no doubt appeal to the student as being straightforward and simple—*if the original assumption as to the width of the transposition rectangle is correct.* But, unfortunately, there is no way of knowing whether such an original assumption is correct until solution is well under way. In practice, of course, what might be done within a well-organized cryptanalytic unit would be to divide up the work among the individuals constituting the.unit, each being assigned one or more specific hypotheses to try out with respect to width of rectangle. Then one of these individuals would find the correct width and he would be joined by the others as soon as an entering wedge had been found in this way. Of, if the cryptanalyst is working alone, he must try out successive hypotheses as to width of rectangle until he hits upon the correct one. In making these hypotheses he must be guided by previous experience with enemy correspondence, which may afford clues as to minimum and maximum widths of rectangles.

b. However, there is another method of attack which does not necessitate making any definite initial assumptions with respect to the width of the transposition rectangle. This method is a modification of the method set forth in the preceding paragraph. The text of the cryptogram is written out columnwise on cross-section paper, every fifth letter being numbered for purposes of reference. Plenty of space is left between the columns, and about 10 or 15 letters at the bottom of each column are repeated at the top of the next column so that at any point in the transcription there will be in a single unbroken string at least one complete column of letters from the transposition rectangle. Then a section of consecutive letters of text is written on a separate strip of cross-section paper, columnwise of course, and by juxtaposing this strip against the whole text, sliding it to various points of coincidence against the text, an attempt is made to find that position in which the best digraphs are formed of the letters on the movable strip and the fixed sequence. Of course, if there is a Q in the cryptogram, the sliding-strip section is made to contain this letter, and the strip is then placed against the text where a U is found, so as to form the digraph QU. The digraphs formed above and below the QU are then studied; possibly a written record is made of the digraphs found. Then the same thing is done with the Q and all other U's in the text, to insure that a correct start is made. It is this initial step which is likely to give the most difficulty (if there is anything difficult at all in the procedure) and it is important that it be correct. If this first step is easy, then solution follows quite rapidly; if the cryptanalyst is unlucky and makes several false starts, the process is likely to be a slow one. In choosing from among several possible juxtapositions it may be advisable to calculate the probability value of each possibility by adding the logarithms of the frequency values of the digraphs, as explained in paragraph 7*g*. In the absence of any Q's in the text, recourse must be had to the formation of other probable digraphs, based upon the presence of certain other telltale low-frequency letters, such as C, H, J, K, V, and X. The cryptanalyst is fortunate if there are two or three of these low-frequency letters close to one another in a series of letters, for in this case he can search for a place where there are high-frequency letters (in a corresponding sequence) that might be combined with them. For example, suppose that a text shows a sequence . . . V E H H K . . . . A sequence such as . . . A R T C C . . . would be excellent to try, for it will yield the digrahhs AV, RE, TH, CH, CK. Or if there is a long sequence of consonants, the cryptanalyst should look for a correspondingly long sequence of vowels, since these make the best combinations and are therefore most probable. For these reasons it pays to study the text quite carefully before choosing a starting point, to find all such peculiar sequences as might be useful in affording a good point of departure. It should also be noted that there are at least two correct positions at which the sliding strip can be juxtaposed against the text, since in the enciphering rectangle the letters in one column form digraphs with the letters in the column not only on the right but also on the left. In the absence of any Q's, or other low-frequency letters suitable for a point of departure, the very first 20 or 25 letters of the cryptogram may be used as the starting point, since these letters come from column 1 of the transposition rectangle and therefore there is no uncertainty at least as to the letter which is at the top of that column; or, the last 20 or 25 letters of the cryptogram may be used as the starting point, since these letters come from the last-numbered column of the rectangle and therefore there is no uncertainty at least as to the letter which is at the bottom of that column.

c. Suppose that a good initial juxtaposition has been found for the portion of the text that has been written on the sliding strip, and that a series of excellent digraphs has been brought to light. The next step is, of course, to add to these digraphs on either side by finding sections of text that will yield "good" trigraphs and tetragraphs. For example, suppose that the initial juxtaposition has yielded what is shown in figure 14. The digraph P R suggests that it must be followed by a vowel, preferably E, A, or O; the digraph A V might be part of the word CAVALRY, in which case it will be followed by A; the digraph C R suggests that it might be followed by the vowel A or E. A place is therefore sought, in the rest of the text, where there is a sequence of the letters here desired, *and, of course, at the proper intervals.* Suppose such a sequence is found and yields what is shown in figure 15. The skeletons of words are now beginning to appear. Assuming that A V A is indeed part of the word CAVALRY, there should be an L to follow it; the trigraph T I N suggests the termination G; the trigraph Z E R suggests the word ZERO. A section of text is therefore sought, which will have the letters L, G, and O in the order $\mathrm{L}\ \overset{1}{.}\ \overset{2}{.}\ \overset{3}{.}\ \overset{4}{.}\ \overset{5}{.}\ \mathrm{G}\ \overset{1}{.}\ \overset{2}{.}\ \mathrm{O}$. Enough has been shown to demonstrate the procedure. In the course of the work it soon becomes evident where the ends of columns are, because the digraphs above and below the nuclear or "good" portion become "bad" quite suddenly, just as soon as letters belonging to nonadjacent columns in the original rectangle are brought together. For example, in figure 15 it is observed that the trigraph at the top, R R S, is highly improbable, as is likewise the trigraph at the bottom, E A O. This suggests that these letters have been brought together erroneously, that is, that they do not belong in adjacent columns in the enciphering rectangle. If this is true then the "good" portion is composed of the 13 letters between these two extremities and therefore the columns are about 13 letters long. Additional work will soon show exactly how long each column really is, and when this has been ascertained the problem has been practically completed, since at the same time that this becomes evident the sequence of columns has also become evident.

.	.
.	.
R	R
N	A
P	R
T	O
A	V
R	E
T	H
C	H
C	K
I	L
T	I
C	R
B	E
Z	E
E	A
.	.
.	.

FIGURE 14.

.	.	.
.	.	.
R	R	S
N	A	T
P	R	E
T	O	R
A	V	A
R	E	D
T	H	R
C	H	U
C	K	A
I	L	L
T	I	N
C	R	A
B	E	S
Z	E	R
E	A	O
.	.	.
.	.	.

FIGURE 15.

d. An example of solution by this alternative method may be helpful. Using the cryptogram of paragraph 13 as an example, figure 16 shows how the text might be transcribed on a sheet of cross-section paper. Noting that the message contains a Q as the 129th letter, a section of text to include the Q is transcribed on a strip of cross-section paper and this strip is then juxtaposed against the remaining text to bring the Q in front of a U. How many letters should be included in this strip? The message contains 224 letters; if a width of say 10 to 20 columns is assumed, the columns of the rectangle will be about 12 to 22 letters in length. It will be safer to assume a convenient length closer to the maximum than to the minimum; consequently a length of 20 letters will be tentatively assumed. Now the Q may be at the top of a column, at the middle,

O 1	D 31	M 61	F 91	T 121	E 151	A 181	A 211
D	S	S	M	N	A	E	L
N	T	C	B	D	O	N	O
N	E	L	I	A	O	E	D
P 5	O 35	C 65	E 95	H 125	T 155	E 185	I 215
T	I	S	H	I	H	M	R
I	T	O	B	R	E	Y	O
R	D	E	V	H	A	T	R
N	M	F	T	Q	C	F	Y
T 10	A 40	C 70	B 100	I 130	N 160	I 190	N 220
D	N	F	E	E	S	T	N
T	L	F	S	T	H	A	R
U	N	T	R	E	E	N	G
R	O	E	S	T	E	L	Y
O 15	E 45	E 75	Y 105	N 135	V 165	N 195	
E	B	E	L	O	T	N	
X	O	M	X	T	R	U	
A	U	I	R	R	E	A	
L	H	A	O	A	S	C	
N 20	E 50	I 80	R 110	A 140	R 170	L 200	
I	N	T	U	V	A	R	
E	L	E	M	R	I	E	
T	E	A	E	I	I	N	
G	S	I	T	R	E	R	
N 25	E 55	J 85	Y 115	S 145	A 175	T 205	
W	A	N	O	T	T	R	
T	A	S	I	G	E	A	
T	C	O	K	S	E	T	
M	T	I	N	E	A	S	
E 30	R 60	V 90	K 120	F 150	L 180	O 210	
D	M	F	T	E	A	A	
S	S	M	N	A	E	L	
T	C	B	D	O	N	O	
E	L	I	A	O	E	D	
O 35	C 65	E 95	H 125	T 155	E 185	I 215	
I	S	H	I	H	M	R	
T	O	B	R	E	Y	O	
D	E	V	H	A	T	R	
M	F	T	Q	C	F	Y	
A 40	C 70	B 100	I 130	N 160	I 190	N 220	

FIGURE 16.

		1			2			3			4		
		O			OT	28		OF	91		OE	177	
	110	R		110	RM		110	RM		110	RE		
		U			UE	30		UB			UA		
		M			MD			MI			ML	180	
		E			ES			EE	95		EA		
		T			TT			TH			TE		
	115	Y		115	YE		115	YB		115	YN		
		O			OO	35		OV			OE		
		IO	1		II			IT			IE	185	
		KD			KT			KB	100		KM		
		NN			ND			NE			NY		
	120	KN		120	KM		120	KS		120	KT		
		TP	5		TA	40		TR			TF		
		NT			NN			NS			NI	190	
		DI			DL			DY	105		DT		
		AR			AN			AL			AA		
	125	HN		125	HO		125	HX		125	HN		
		IT	10		IE	45		IR			IL		
		RD			RB			RO			RN	195	
		HT			HO			HR	110		HN		
→		QU			QU			QU			QU		←
	130	IR		130	IH		130	IM		130	IA		
		EO	15		EE	50		EE			EC		
		TE			TN			TT			TL	200	
		EX			EL			EY	115		ER		
		TA			TE			TO			TE		
	135	NL		135	NS		135	NI		135	NN		
		ON	20		OE	55		OK			OR		
		TI			TA			TN			TT	205	
		RE			RA			RK	120		RR		
		AT			AC			AT			AA		
	140	AG		140	AT		140	AN		140	AT		
		VN	25		VR	60		VD			VS		
		RW			RM			RA			RO	210	
		IT			IS			IH	125		IA		
		RT			RC			RI			RL		
	145	SM		145	SL		145	SR		145	SO		
		TE	30		TC	65		TH			TD		
		GD			GS			GQ			GI	215	
		SS			SO			SI	130		SR		
		ET			EE			ET			EO		
		1			2			3			4		

FIGURE 17-A.

or at the bottom—there is no way of telling at this point. Hence, to make sure that nothing is overlooked, suppose a section of 41 letters is taken, with the Q at the center. There are 4 U's in the message, and 4 trials are to be made. The results are as indicated in figure 17–A. Examining combination 1 in figure 17–A, the digraphs formed both above and below the Q U are not at all bad. In fact, not one of those above the Q U is impossible and the same is true of those below the Q U until the digraph V N is reached. Hence, combination 1 is possible. As for combination 2, this at once appears to be bad. Digraphs such as I I, and I H are highly improbable, and this combination may be discarded with safety. Combination 3 is possible from the top digraph, O F, to the 12th digraph below the Q U, although the digraph H X looks very bad. However, the X might be a sentence separator, so that this combination cannot be discarded. Combination 4 looks very improbable, with the digraph H N occurring twice, and other equally bad digraphs showing. Of the four possibilities then, combinations 2 and 4 are discarded, leaving 1 and 3 for further study. It is very difficult to choose between these two possibilities. All the digraphs in combination 1 down to digraph V N are possible; many of them are excellent. As for combination 3, all the digraphs down to V D are also possible and many of them are excellent. There does not seem to be much use to add the frequency values of the digraphs (or logarithms thereof) in each combination because it is hard to know with what digraphs to begin or end, although as a last resort this could of course be done. *However, perhaps it is not essential that a choice be made at once; possibly further work along the lines now to be demonstrated will show which combination is correct.*

		SOF
		ERM
		AUB
		AMI
		CEE
		TTH
		RYB
		MOV
	SIO	SIT
	CKD	CKB
	LNN	LNE
	CKN	CKS
	STP	STR
	ONT	ONS
	EDI	EDY
	FAR	FAL
	CHN	CHX
	FIT	FIR
	FRD	FRO
	THT	THR
➔	EQU	EQU
	EIR	EIM
	EEO	EEE
	MTE	MTT
	IEX	IEY
	ATA	ATO
	INL	INI
	TON	TOK
	ETI	ETN
	ARE	ARK
	IAT	IAT
	JAG	JAN
	1	3

FIGURE 17-B.

Noting the 2 K's (in the digraphs K B and K S) among the combinations before the Q, assume that these K's are parts of the digraph C K. Is there a sequence C . C in the text? There is but one such place, at the 63d letter. Suppose the corresponding section is placed in front of the combinations 1 and 3 of figure 17, as shown in figure 17–B. It immediately becomes evident that combination 3 is the correct one, for note the excellent trigraphs it gives, as compared with those in combination 1. Also note that the second trigraph below the E Q U in combination 3 consists of 3 E's, indicating that the end of the columns has been reached just before this trigraph. As for the top trigraphs of figure 17–B they are good all the way up. But now the skeletons of words are beginning to appear. The T H R immediately above the E Q U suggests either THREE or THROUGH; the F R O above the T H R suggests FROM or FRONT. Suppose the word REQUEST is assumed for the E Q U, and the word THREE is assumed for the T H R above it. This requires a section with 2 E's in succession.

SOFV	
ERMT	
AUBR	
AMIE	
CEES	
TTHR	
RYBA	
MOVI	
SITI	
CKBE	
LNEA	
CKST	
STRE	
ONSE	
EDYA	
FALL	
CHXA	
FIRE	
FRON	
THRE	
EQUE	←
EIMM	
EEEY	

FIGURE 17-C.

e. There are several such places in the text, and further limitation is advisable. The 8th trigraph from the top is certainly suggestive of the word MOVING, which requires an I to follow the V. Is there a place in the text where an I occurs 12 letters before a succession of two E's? There is one such place, and the corresponding section is juxtaposed at the proper place, yielding what is shown in figure 17–C. The upper and lower limits of the columns are now fairly definite and are marked by the horizontal bars; tetragraphs E E E Y at the bottom and A M I E at the top are very improbable. The tetragraph C E E S below the top bar is possible, because it may represent the end of a word like FORCE followed by the beginning of the word ESTIMATED; the tetragraph above the bottom bar suggests a word ending in E followed by the word IMMEDIATE. It seems hardly necessary to continue with the demonstration; in a few moments the entire diagram is reconstructed and yields the solution. During this process, as soon as a section of text in figure 16 has been used it is crossed off, so as to prevent its letters from being considered as further possibilities for addition to the reconstruction diagram. Thus, as the work progresses the number of available sections becomes progressively less, and the choice for successive sections for addition to the diagram becomes a quite easy matter.

f. When two or three operators are assigned to work upon a cryptogram by this method, solution can be reached in a very short space of time, especially if each one of the operators takes a different point of attack. After a few minutes the fragments of texts obtained may be assimilated into one message which is then completed very speedily.

g. This and the next four subparagraphs will be devoted to some remarks of a general nature concerning columnar transposition of the foregoing type. The degree of cryptographic security afforded by simple columnar transposition methods, especially when incompletely-filled matrices are employed, is considerably increased if some of the cells of the matrix are occupied by nulls instead of significant letters. If nulls are employed judiciously their presence serves to confuse the cryptanalyst by introducing unusual digraphs, trigraphs, and polygraphs which may lead him to discard correct combinations of columns in the anagramming process and thus retard solution. Obviously, the use of low-frequency letters such as J, Q, X, or Z as nulls does not commend itself for this purpose, as such letters would not only distort the normal frequency distribution and thus give clues to the presence of nulls, but also they would be quickly "spotted" in the anagramming process.

h. Another subterfuge, and a good one, to put stumbling blocks in the way of a quick solution is to leave "blanks" within the transposition matrix, that is, certain cells are left unoccupied by letters of the text. If only a few cells distributed irregularly within the columns of the transposition matrix are designated as blanks, the disturbing effect upon the anagramming process is quite marked. This more or less effectively hinders the cryptanalyst in his attempts to ascertain the lengths of the columns and considerably increases the difficulty of the anagramming process.

i. In order to fix definitely the positions of the nulls or of the blanks in the transposition matrix, definite prearrangements between correspondents are necessary. These may be in the nature of "forms" outlining the matrix, showing the number of columns and the positions of the cells to be occupied by nulls, or of the cells to be left vacant in the inscription process; or the positions of these cells may be derived from the elements of the transposition key itself. If "forms" are employed, they may be used with varying transposition keys, so that even though there may be only relatively few different forms, the use of varying transposition keys serves to increase cryptographic security to a rather marked degree.

j. If nulls, or blanks, or both, are distributed irregularly but symmetrically throughout the transposition matrix (as, for example, blanks are distributed in cross-word puzzles) solution of single messages produced by simple keyed-columnar transposition from such a matrix becomes an extremely difficult if not impossible problem. Naturally, if nulls and blanks are distributed irregularly and assymmetrically the matter becomes hopeless, as far as a single message is concerned.

k. Of course, if several messages of identical length and in the same key are available for superimposition, the presence of the nulls or blanks then makes little difference, because the *general solution* to be explained in a subsequent paragraph (par. 26) can be applied. Or if messages with similar beginnings or similar endings are available, solution is facilitated here as in the simpler case where nulls or blanks are not employed, as will be explained in subsequent paragraphs (pars. 23–24). Considerations of space prevent going into detail in the solution of an example, and the student should undertake a study of these cases for himself.

16. The C→P and the P→C sequences.—*a.* Two numerical sequences which constitute the bases for several very important cryptanalytic operations and procedures in the solution of transposition ciphers may be derived from, and are applicable to, most ciphers of this class. They are as follows:

(1) A sequence the successive terms of which indicate the position numbers that the successive letters of the plain text occupy in the cipher text. This sequence will hereinafter be designated the *plain→cipher sequence*, or *P→C sequence*.

(2) A sequence the successive terms of which indicate the position numbers that the successive letters of the cipher text occupy in the plain text. This sequence will hereinafter be designated the *cipher→plain sequence*, or *C→P sequence*.

b. These two sequences are obviously related, one being the *inverse* or indexed version of the other. Given one of the sequences, the other can be derived from it by the simple operation of indexing, in a normal sequence, the positions occupied by the numbers constituting the sequence on hand. An example will be given presently.

c. Note the encipherment shown in Figure 18–A.

T	R	A	N	S	P	O	S	I	T	I	O	N
12	9	1	4	10	8	6	11	2	13	3	7	5
T	H	E	Q	U	I	C	K	B	R	O	W	N
F	O	X	J	U	M	P	S	O	V	E	R	T
H	E	L	A	Z	Y	D	O	G				

Term No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Cipher	E	X	L	B	O	G	O	E	Q	J	A	N	T	C	P	D	W	R	I	M	Y	H	O	E	U	U	Z	K	S	O	T	F	H	R	V

FIGURE 18–A.

Now, if, instead of letters, the successive numbers 1, 2, 3, . . . are inscribed in the cells of the matrix, in normal order of writing, the "cipher text" becomes the P→C sequence and is as follows:

12	9	1	4	10	8	6	11	2	13	3	7	5
01	02	03	04	05	06	07	08	09	10	11	12	13
14	15	16	17	18	19	20	21	22	23	24	25	26
27	28	29	30	31	32	33	34	35				

Term number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
P→C equences	03	16	29	09	22	35	11	24	04	17	30	13	26	07	20	33	12	25	06	19

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
32	02	15	28	05	18	31	08	21	34	01	14	27	10	23

FIGURE 18–B.

The student may easily verify that the P→C sequence is what it purports to be by noting that, according to it, the 1st letter of the plain text of the illustrative message, T_p, becomes the 31st letter of the cipher text (since the number 01 occupies position 31 in the P→C sequence shown above), and that in the cryptogram the 31st letter is T_c; the 2d letter of the plain text, H_p, becomes the 22d letter of the cipher text and that in the cryptogram the 22d letter is H_c; and so on. In connection with the P→C sequence, it is to be noted that successive terms in the sequence, in the case of single transposition, show a constant difference except when passing from a greater to a smaller number, which happens every time a transition is made from a term applying to the bottom element of one column to a term applying to the top element of the next column. For example, in the case of the 1st three terms in the sequence: 16−03=13; 29−16=13. However, in the case of the 3d term of the sequence (29) and the 4th (09) the passage is from a greater to a smaller number and the constant difference, 13, no longer is evident. The cause of the constant difference is, of course, obvious and follows directly from the mechanics of the transposition system itself. The point to be specially noted is that the existence of such a constant difference (with the exceptions noted above) may be taken as one of the identifying characteristics of single columnar transposition; double columnar transposition or other types of complex transposition will show no such constant difference throughout the P→C sequence.

d. Given the P→C sequence in subparagraph *c*, its inverse, the C→P sequence is established merely by preparing an indexed version of the former. Thus:

Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
P→C sequence	03	16	29	09	22	35	11	24	04	17	30	13	26	07	20	33	12	25	06	19
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35					
	32	02	15	28	05	18	31	08	21	34	01	14	27	10	23					

Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
C→P sequence	31	22	01	09	25	19	14	28	04	34	07	17	12	32	23	02	10	26	20	15
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35					
	29	05	35	08	18	13	33	24	03	11	27	21	16	30	06					

e. The C→P sequence can also be produced in another way. Suppose that numbers are inscribed in the cells of the transposition matrix, not in the normal manner of writing from left to right and form the top downward, but according to the route followed in *transcribing* the numbers to form the "cipher text," that is, in key-number order in the columns of the matrix. Thus:

12	9	1	4	10	8	6	11	2	13	3	7	5
31	22	01	09	25	19	14	28	04	34	07	17	12
32	23	02	10	26	20	15	29	05	35	08	18	13
33	24	03	11	27	21	16	30	06				

FIGURE 18–C.

If these numbers are now transcribed according to the normal manner of writing (from left to right and from the top downward), the sequence produced is 31 22 01 09 25 . . ., which coincides with the C→P sequence shown in subparagraph *d* above, which in turn was derived from the P→C sequence.

Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
C→P sequence	31	22	01	09	25	19	14	28	04	34	07	17	12	32	23	02	10	26	20	15
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35					
	29	05	35	08	18	13	33	24	03	11	27	21	16	30	06					

f. The C→P sequence may also be called the *anagram sequence* because it can be established as a result of a solution accomplished by anagramming superimposed messages of identical length. It is clear that what is accomplished in such a solution is to rearrange the letters of the cipher text to bring them back into their original order in the cipher text, that is, the solution involves a C→P conversion.

g. The P→C sequence is called by a recent French author the *κρ* sequence (from the Greek word *kryptos*) because it gives the order of the plain-text letters as they occur in the cryptogram. The P→C sequence is also termed the *encipher sequence* by another writer, and still another has called it the *transposition sequence*. The present author believes that the terminology adopted herein, *viz*, P→C sequence and C→P sequence, is less confusing and serves more accurately to identify or characterize these sequences than the other designations herein indicated.

h. The *term number* is useful merely to facilitate finding and referring to specific terms or numbers in a sequence, whether the latter be a C→P or a P→C sequence. The number simply indicates the locus or position a term occupies in the sequence. In connection with a plain-text message the consecutive term numbers 1, 2, 3, . . . may be used as loci for the successive letters of the message; in connection with a cryptogram the consecutive term numbers 1, 2, 3, . . . may be used as loci for the successive letters of the cipher text.

i. In single, keyed-columnar transposition an interesting relationship exists between sections of the C→P sequence. Consider the C→P sequence given in subparagraph *d* above, and note that by adding the integer 1 to the successive numbers thereof, sections of the original sequence show certain identities with sections in C→P sequence +1. Thus:

Term number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
C→P sequence	31	22	01	09	25	19	14	28	04	34	07	17	12	32	23	02	10	26	20	15	29	05
C→P sequence +1	32	23	02	10	26	20	15	29	05	35	08	18	13	33	24	03	11	27	21	16	30	06

Term number	23	24	25	26	27	28	29	30	31	32	33	34	35
C→P sequence	35	08	18	13	33	24	03	11	27	21	16	30	06
C→P sequence +1	36	09	19	14	34	25	04	12	28	22	17	31	07

In fact, if the successive numbers of the C→P sequence are set down in rows to produce sequent numbers in *columns*, the following interesting diagram is obtained:

	31	22	01	09	25	19	14	28	04	34	07	17	12
+1	32	23	02	10	26	20	15	29	05	35	08	18	13
+1	33	24	03	11	27	21	16	30	06				

FIGURE 18-D.

Reference to figure 18–C will show the identity of this diagram with that figure. Such an arrangement of course indicates at once the number of columns in the transposition rectangle, from which it follows that if the C→P sequence is available it is an easy matter to establish the outlines of the transposition matrix. The phenomena dealt with in this subparagraph are but a reflection of those discussed in subparagraph *c* above.

j. The phenomena just indicated may, however, be employed to advantage in another manner in the solution of an unknown example. Referring to the illustrative cryptogram in subparagraph *c* above, suppose that the cryptanalyst has reason to suspect the presence of the probable word QUICK. The letters necessary to produce this word (and their term numbers in the cryptogram) are as follows:

9	25 26	19	14	28
Q	U	I	C	K

The sequence $9\text{–}{}^{25}_{26}\text{–}19\text{–}14\text{–}28$ now constitutes a portion of the C→P sequence. Adding the integer 1 successively to these C→P numbers, let the corresponding letters be set down alongside the numbers. Thus:

Base	9	25 26	19	14	28	=	Q	U	I	C	K
Derivative 1	10	26 27	20	15	29	=	J	U Z	M	P	S
Derivative 2	11	27 28	21	16	30	=	A	Z K	Y	D	O
Derivative 3	12	28 29	22	17	31	=	N	K S	H	W	T

Here it will be seen that portions of "good" plain text become manifest, viz, JUMPS and AZYDO. The 3d derivative no longer is "good" because the rectangle has but 3 rows and consequently only the 1st and 2d derivatives from the "base" are valid. It is obvious that the foregoing method of deriving plain-text sections from a correct probable word offers considerable possibilities as a cryptanalytic tool, especially in the case of matrices with more than 2 or 3 rows. If sections of text can be reconstructed in this manner and then combined in proper sequence the reconstruction of the complete matrix and the transposition key is a relatively simple matter. The application of the foregoing principle to the solution of unknown examples is, of course, obvious.

k. There is also an interesting relationship between the sections of the P→C sequence for a cryptogram, though it is somewhat different from that discussed is subparagraphs *i* and *j* in the case of the C→P sequence. Consider the P→C sequence set forth in subparagraph *d* above and note how, by adding the integer 1 to the successive numbers, sections of the P→C sequence become identical with sections of the P→C+1 sequence. Thus:

P→C sequence..........

$$\underbrace{03\ 16\ 29}_{1}\ \underbrace{09\ 22\ 35}_{2}\ \underbrace{11\ 24}_{3}\ \underbrace{04\ 17\ 30}_{4}\ \underbrace{13\ 26\ 07}_{5}\ \underbrace{20\ 33}_{6}\ \underbrace{12\ 25}_{7}$$

$$\underbrace{06\ 19\ 32}_{8}\ \underbrace{02\ 15\ 28}_{9}\ \underbrace{05\ 18\ 31}_{10}\ \underbrace{08\ 21\ 34}_{11}\ 01\ \underbrace{14\ 27}_{12}\ \underbrace{10\ 23}_{13}$$

P→C sequence+1......

$$\underbrace{04\ 17\ 30}_{1=4}\ \underbrace{10\ 23}_{2=13}\ 36\ \underbrace{12\ 25}_{3=7}\ \underbrace{05\ 18\ 31}_{4=10}\ \underbrace{14\ 27}_{5=12}\ \underbrace{08\ 21\ 34}_{6=11}\ \underbrace{13\ 26\ 07}_{7=5}$$

$$\underbrace{20\ 33}_{8=6}\ \underbrace{03\ 16\ 29}_{9=1}\ \underbrace{06\ 19\ 32}_{10=8}\ \underbrace{09\ 22\ 35}_{11=2}\ \underbrace{02\ 15\ 28}_{12=9}\ \underbrace{11\ 24}_{13=3}$$

The equivalencies between identities, as indicated above, indicate not only that the enciphering matrix has 13 columns, but also they may be used to establish the actual transposition key or at least a cyclic permutation of the key, by constructing a chain of equivalents. Thus:

1=4, 4=10, 10=8, 8=6, 6=11, 11=2, 2=13, 13=3, 3=7, 7=5, 5=12, 12=9, 9=1

This yields, by eliminating the term common to successive equivalents, the following chain or transposition key:

1 4 10 8 6 11 2 13 3 7 5 12 9

Reference to figure 18–A will show that the foregoing key is a cyclic permutation of the actual key.

l. There remain only some minor remarks which, being of a general nature arising from the mechanics of simple keyed-columnar transposition, are worth noting. They are discussed in the subsequent two subparagraphs.

m. An appreciation of the difficulties introduced by employing only incompletely-filled rectangles indicates that it would be very useful if there were some method whereby in the initial stages of solution the cipher text could be divided up correctly into its component long and short columns, for the subsequent steps of rearranging the columns by the anagramming principle are quite simple. If, for example, there were some feature which provided a means of ascertaining when in encipherment a transit was made from the bottom of one column to the top of the next column, then the location of these transition points or "breaks" would obviously permit of breaking up the cipher text into its correct long and short columns. In later studies cases of this kind will be encountered.

n. It is useful sometimes to be able to ascertain just where breaks cannot occur, that is, where a passage from the bottom of one column to the top of the next one cannot occur in the cipher text, for this will limit the field for experiment. A consideration of the mechanics of the system will afford an excellent clue to the fact that this determination is easy to make. In any transposition rectangle involving simple keyed-columnar transposition the interval, in the cipher text, between two consecutive letters which are in the same row in the matrix is the sum of a multiple of the length of the short columns and a multiple of the length of the long columns. For example, consider the adjacent letters C K in the plain-text rectangle in figure 18–A. In the cipher, C_c is the 14th letter, K_c is the 28th and the interval is $28-14=14$. The message has a total of 35 letters and the matrix has 13 columns, 9 long ones of 3 letters and 4 short ones of 2 letters. An interval of 14 can therefore be brought about in only one way: $(4\times3)+(1\times2)=14$, which means that 4 long columns and 1 short one intervene between the C and the K in the plain text, and that the key numbers applicable to the two columns are 5 apart in sequence, that is, if the column in which C is located has key number 1, the column next to it on the right is 6, or if the former is 2 the latter is 7, and so on. Reference to figure 18–A will show that these deductions are correct and that the key numbers involved are 6–11. However, a more general treatment is possible. Given a cryptogram of 26 letters and an assumed width of 6 columns, for example, the matrix can have only 2 columns of 5 letters and 4 columns of 4 letters. Setting down the multiples of the two lengths in tabular form, for convenience, the following is obtained:

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26				

	5	4
0 multiple	0	0
1st multiple	5	4
2d multiple	10	8
3d multiple	—	12
4th multiple	—	16

All the possible positions of breaks in the cipher text, that is, transits from the bottom of one column to the top of the next column, may now readily be ascertained by finding the totals resulting from making all the possible combinations of the indicated multiples taken in pairs. It is convenient to draw up a table to show directly the sums of the combinations. Thus:

		0	1	2	3	4	No. of short columns
		0	4	8	12	16	Length in letters
No. of long columns	Length in letters						
0	0	0	4	8	12	16	
1	5	5	9	13	17	21	
2	10	10	14	18	22	26	

If, now, diagonal lines are drawn from the lower left-hand corner to the upper right-hand corner of the diagram, the locations of all possible breaks are given. Thus, there can be a break between the 4th and 5th letters (passing from a short column to the next column, which may be long or short, of course); there cannot be a break between the 5th and 6th letters, nor between the 6th and 7th, nor between the 7th and 8th; there can be a break between the 8th and 9th, as well as between the 9th and 10th, but not between the 10th and 11th, and so on. Suppose that for one reason or another the cryptanalyst has good reason to suspect that a break occurs immediately after the 13th letter. This means that there are 2 short columns (of 4 letters) and 1 long column (of 5 letters) up to that break. The diagram shows that there remain only 2 short columns and 1 long one, and the only breaks that are possible beyond the one at the 13th letter are: Between the 17th and 18th, or between the 21st and 22d letters.

o. The importance of the various principles set forth in this paragraph will become evident as the student progresses in his studies of transposition ciphers.

SECTION IV

OPPORTUNITIES AFFORDED BY STUDYING ERRORS AND BLUNDERS MADE BY ENEMY CRYPTOGRAPHERS

Paragraph
Importance of the study of errors and blunders in early work upon an unknown system........................ 17
Significance of terms "special solution" and "general solution"........................ 18
Examples to be studied........................ 19

17. Importance of the study of errors and blunders in early work upon an unknown system.—*a.* Blunders and mistakes made by cryptographic clerks in the execution of cryptographic clerks in the execution of cryptographic instructions should be rare in a well-trained and well-disciplined cryptographic service. Nevertheless, blunders and mistakes are committed despite all that can be done to prevent their occurrence. Especially in the excitement prior to or during an important action or movement do such instances take place and these afford golden opportunities for the enemy cryptanalytic service. This situation exists in respect to all types of cryptographic systems and no cryptanalytic instruction would be complete if cognizance were not taken of the advantages which may be reaped from the blunders, the mistakes, and, occasionally, the downright ineptitude of the adversary's cryptographers.

b. Practically every cryptographic system affords opportunities for the commission of errors in its application, and each system more or less presents a separate case. That is, the errors which may be made in one type of cryptographic system may be peculiar to that type alone and to no other type; hence, the astute cryptanalyst is constantly on the lookout for instances of cryptograms containing the specific type of error by which that system is handicapped. Furthermore, the general types of blunders or errors that may be committed are nearly as numerous as are the general types of cryptographic systems, so that no complete list of such as may be encountered in practice can be drawn up.

c. After the cryptanalyst has by painstaking and more or less arduous labors solved a system and has become thoroughly familiar with its mechanics, he should carefully review the details of the mechanics to learn what things can go wrong, what sorts of mistakes the enemy cryptographic personnel are likely to make, and *then study the external manifestations of these aberrations so that he may be able to recognize instances of their occurrence in subsequent cryptograms.* This sort of study has no value in itself particularly; its importance lies in the fact that the effects of erroneous treatment may lead to very rapid solution or to quick recovery of keys to subsequent messages.

d. When an unknown system is under investigation and the cryptanalyst is striving to ascertain just how it operates (which is often the most difficult step in solution), a study of the cryptograms representing corrections to previous messages containing errors is a most fruitful source of data. Indeed, at times this sort of intensive study will yield clues for solving a system which might otherwise resist all efforts to break it down for a very long time.

18. Significance of terms "special solution" and "general solution."—*a.* Now the importance of the comments made in the foregoing paragraph will be clear if it is noted that a study of the blunders and errors often leads to the elaboration of methods for the rapid breaking down of cryptographic systems. But it must also be realized that in some cases no blunders or errors are essential to a rapid solution of the type alluded to above. Sometimes the very mechanics of the system itself are such that unavoidable or unpredictable circumstances arise so that special solutions become possible. The latter term calls for a bit of explanation.

b. When the circumstances surrounding a specific cryptogram or set of cryptograms are such as to present peculiar or unusual conditions that make a solution possible when in the absence of these conditions solution is either impossible or improbable, the methods employed in reaching a solution in such cases constitute what is commonly termed a *special solution*. Some examples will be demonstrated very soon. Systems of which this may be true are, of course, cryptographically weak but it may be observed that it is perhaps impossible to devise a system which may be considered to be absolutely free from this source of weakness.

c. The advantages of a special solution for any type of cryptographic system are, as a rule, two in number. First, it often makes a solution possible when otherwise this might not be the case. Secondly, if often affords a method of achieving a very rapid solution in the case of a problem which otherwise might require a long time. But a special solution presents one basic disadvantage: It is by its very nature dependent upon the existence of unusual circumstances; in other words, upon chance or good fortune bringing about a set of circumstances favorable for a solution. When these unusual conditions or circumstances do not obtain, then solution may be impossible. Therefore, it is desirable to have, if possible, for every type of system a more or less *general solution* which may be applied in the absence of the unusual conditions necessary for the application of a special solution. In other words, a general solution in cryptanalysis implies a method or procedure which if applied in ordinary cases and under normal conditions will yield the solution. However, the term *general solution* in cryptanalysis must not be taken too literally. The situation in cryptanalysis is not exactly analogous to that which obtains in the field of pure mathematics, for the circumstances are often quite different in the two sciences. A general solution in mathematics is expected to, and will, solve every case that falls within its province; a general solution in cryptanalysis is likewise intended to solve every case that falls within its province but this usually partakes more of the nature of a prayer or hope than an expectation. Much depends upon the amount of traffic available for study, the length of individual cryptograms, and the indefinable element called luck, that is, a set of fortuitous circumstances which happen to make a solution easy or difficult, such as the presence of many or exceptionally long repetitions, etc. Furthermore, whereas in mathematics a general solution prescribes the exact steps to be followed in arriving at the solution and the latter can be applied in all instances without variation or deviation from a fixed procedure, in cryptanalysis a general solution merely outlines a broad path that may be followed in order to arrive at a solution. Application of a general solution in cryptanalysis in specific instances may involve minor detours to circumvent unexpected obstacles, or it may involve quite large changes or modifications in the general procedure.

19. Examples to be studied.—*a.* As stated above in paragraph 17, a complete list of the specific blunders that cryptographic clerks are prone to perpetrate cannot be drawn up. Certain of them may be described in general terms and examples given of some which have already been encountered in this and in preceding texts. Commonly it is the case that these blunders do not become evident until two or more cryptograms are available for comparison. One of the most frequent sources of circumstances leading to the transmission of cryptograms affording rich material for cryptanalytic comparison is the following: A cryptographic clerk prepares a cryptogram, in the course of which he makes a mistake of such a nature as to render the cryptogram difficult or impossible to decipher by the cryptographic clerk serving the addressee. A request for repetition ensues, whereupon the enciphering clerk reexamines his original work and finds that he has made a mistake. He then commits the grave blunder of reenciphering the identical message (without paraphrasing) and transmitting what to the enemy cryptanalysts is obviously a second version of the original message. The consequences are often fatal to cryptographic security. The least that can happen is that the key for this particular message may be disclosed very quickly; more serious, the basic or primary elements for the entire day's traffic may be wrested from the blunder; but most serious are the consequences if it happens that the blunder has been committed immediately or soon after a new cryptographic system has been instituted and the enemy cryptanalysts are exerting strenuous efforts to learn its mechanics, for then is when the information to be gained is most valuable.

b. In his previous studies the student has observed the many opportunities for quick cryptanalytic success afforded by enemy addiction to the use of stereotypic phraseology, especially at the beginnings and endings of messages. Stereotypic phraseology affords even more golden opportunities for cryptanalytic success in the case of transposition systems than it does in the case of substitution systems.

c. In the next few paragraphs some specific examples of the consequences of cryptographic blunders and ineptitude in the case of transposition systems will be studied. These are intended to give the student some idea of the far-reaching effects such studies may have. It is important that he grasp the fundamental principles, for they will enable him to develop for himself the methods that he may find necessary in practical work. Incidentally, it may be added that the student should not get the idea that these instances are purely theoretical. It is sometimes almost unbelievable that cryptographic clerks with any common sense would perpetrate the stupid blunders that they do occasionally commit.

SECTION V

SPECIAL SOLUTIONS FOR TRANSPOSITION CIPHERS

Paragraph
Solution when the beginning or end of the plain text is known........ 20
The case of an omitted column........ 21
The case of an interchanged pair of columns........ 22
Messages with similar beginnings........ 23
Messages with similar endings........ 24
The solution of a single message containing a long repetition........ 25
Solution when several cryptograms of identical length and in the same key are available........ 26
Reconstruction of the keys in double transposition........ 27
Special cases of solution of double transposition ciphers........ 28

20. Solution when the beginning or end of the plain text is known.—*a.* It often happens, when correspondents have fallen into the bad habit of sending stereotyped communications, that the beginnings or the endings of messages become so fixed in their form and content that the enemy can with a fair degree of certainty guess what these will be in specific cases. If so, a quick solution can be reached, the key reconstructed for one message, and this will, of course, enable him to read all other messages in the same key. This is particularly true of simple, keyed-columnar-transposition ciphers. It is only necessary that the cryptanalyst cut the text up in such a manner as to bring the letters composing the assumed text all within the same row or rows of the transposition rectangle.

b. Suppose that the enemy frequently uses the introductory expression REFERRING TO YOUR NUMBER. Here is a cryptogram assumed to begin with this phrase:

CRYPTOGRAM

```
IMAOD  RMGRN  ERNIN  TUSFS  DRYEP  BRCFT
OIRNW  TMOIS  OIEGE  DHOPN  CHLFU  ESEPQ
ERIAR  UHIAG  PAUOO  SSSCI  ONRRE  OVOEY
EMEVG  TRIAF  HTEPB  NBTNE  AEETA
```

c. Assuming that previous experience has indicated that the enemy uses keys varying from 10 to 20 letters in length, the arrangement of the letters in the tops of columns under a key length of 10 would be as shown in Fig. 20.

1	2	3	4	5	6	7	8	9	10
R	E	F	E	R	R	I	N	G	T
O	Y	O	U	R	N	U	M	B	E
R									

FIGURE 20.

The first group of the cryptogram begins with I M. The arrangement shown above gives I U as the top of a column; hence a key length of 10 is not correct. A key length of 11 is then tried.

1	2	3	4	5	6	7	8	9	10	11
R	E	F	E	R	R	I	N	G	T	O
Y	O	U	R	N	U	M	B	E	R	

FIGURE 21.

Here a column is headed by I M, so that this is a possible arrangement. If the width of the rectangle is 11, its outlines are as shown in figure 22. There are 5 columns of 11 letters and 6 columns of 10 letters. The text can now be marked off into sections of proper lengths and, moreover, guided by the letters which must be at the heads of columns, the text can be inscribed in the rectangle in key order. For example, column 1 must end with the second group, R M G R N; column 2 therefore begins with E R. There is only one possibility, viz, the fourth column. This is a long column, and must therefore have 11 letters, making column 3 begin with R Y. This definitely fixes the position of the number 3 in the key, and so on. The solution is reached after only a very few moments and is as shown in figure 23.

R	E	F	E	R	R	I	N	G	T	O
Y	O	U	R	N	U	M	B	E	R	
						A				
						O				
						D				
						R				
						M				
						G				
						R				
						N				

Figure 22.

3	9	6	2	4	7	1	11	5	10	8
R	E	F	E	R	R	I	N	G	T	O
Y	O	U	R	N	U	M	B	E	R	S
E	V	E	N	W	H	A	T	D	I	S
P	O	S	I	T	I	O	N	H	A	S
B	E	E	N	M	A	D	E	O	F	C
R	Y	P	T	O	G	R	A	P	H	I
C	E	Q	U	I	P	M	E	N	T	O
F	M	E	S	S	A	G	E	C	E	N
T	E	R	F	O	U	R	T	H	P	R
O	V	I	S	I	O	N	A	L	B	R
I	G	A	D	E						

FIGURE 23.

d. The same general principles, modified to suit the circumstances, may be followed in the case involving known or suspected endings of messages. The probable words are written out according to various assumed key lengths and the superimposed letters falling at the bottoms of columns are sought in the cryptogram.

21. **The case of an omitted column.**—*a*. Sometimes a very careless clerk omits a column in transcribing the text from the enciphering rectangle and fails to check the number of letters in the final cryptogram. Obviously such a cryptogram will be difficult if not impossible to decipher at the other end, and a repetition is requested and sent. If now the identical plain text is enciphered correctly, two cryptograms are at hand for comparison. This will disclose the length of one column, which can be assumed to be either a long one or a short one. The position, in the correct cryptogram, of the column omitted from the incorrect one will often afford direct clues as to the exact dimensions of the enciphering rectangle. For example, suppose the cryptogram in paragraph 20*b* had first been transmitted as follows:

CRYPTOGRAM

```
I M A O D   R M G R N   R Y E P B   R C F T O   I R N W T   M O I S O
I E G E D   H O P N C   H L F U E   S E P Q E   R I A R U   H I A G P
A U O O S   S S C I O   N R R E O   V O E Y E   M E V G T   R I A F H
T E P B N   B T N E A   E E T A
```

b. The column which was omitted is E R N I N T U S F S D, and falls between columns 1 and 3. Since the omitted column contains 11 letters and column 1 contains 10, the dimensions of the rectangle immediately become known. Thus, uncertainties as to the dimensions of the rectangle are dissolved and a large step forward has been made in the solution. Also, the general whereabouts of columns 1 and 2 are now known, since the former is a short one, the latter a long one.

22. The case of an interchanged pair of columns.—*a.* The keying element in the case of columnar transposition is simply a practical means of controlling the order in which the columns of the enciphering rectangle are transcribed in forming the cipher text. Commonly this numerical key is derived from a literal key. Suppose that a cryptographic clerk makes a mistake in the latter step. For example, suppose that the literal key is ADMIRATION and that as a result of a slight relaxation in attention he assigns the number 5 to the letter N and the number 6 to the letter M. A pair of columns will become interchanged as regards their order of selection in the transcription process, and likely as not a repetition will be requested by the addressee. If a second version is sent, enciphered by the correct key, a comparison of the two versions will disclose the width of the enciphering rectangle and possibly the general position (left or right) of the columns that were interchanged.

b. An example will serve to make the matter clear. Assume the two cryptograms to be as follows:

First Version

```
O D N I L   N T T H D   G S O H A   O O Q S G   T E R P S
I N E N E   N F U E H   R W R R I   R A T P E   D E T A N
O O C O O   R O G I O   S
```

Second Version

```
O D N I L   N T T H D   G S O H A   O O Q S G   T E R N F
U E H R W   R P S I N   E N E R I   R A T P E   D E T A N
O O C O O   R O G I O   S
```

c. The two cryptograms are superimposed as shown in figure 24 and their points of similarity and difference noted.

```
First version...  O D N I L N T T H D G S O H A O O Q S G T E R[P S I N E N E]
Second version.   O D N I L N T T H D G S O H A O O Q S G T E R[N F U E H R W
                 [N F U E H R W R]R I R A T P E D E T A N O O C O O R O G I O S
                  R]P S I N E N E]R I R A T P E D E T A N O O C O O R O G I O S
```

Figure 24

d. The two versions are alike except for a pair of interchanged sequences; the bracketed sequence P S I N E N E in the first version is matched by the same sequence in the second version, but at a different position in the message; likewise the bracketed sequence N F U E H R W R in the first version is matched by a similar sequence in the second version, but at a different position in the message. The various deductions which can be made from the situation will now be set forth.

e. One of these sequences contains 7 letters, the other contains 8. It follows that the columns of the enciphering rectangle are probably 7 and 8 letters in length; hence, with 61 letters, the width of the matrix is 8. Since there are 23 letters from the beginning of the messages to the first point of their difference, it follows that there are 2 columns of 8 letters and 1 column of 7 letters involved in this section [$(2\times8)+(1\times7)=23$], and that the error made in encipherment does not involve columns 1, 2, or 3, which are therefore properly placed in the first version. Since the sequences which are interchanged are consecutive in the text it means that the numbers 4 and 5 were interchanged in the key for the first version. Since one of these sequences is of 7 letters, the other of 8 letters, one of the numbers, 4 or 5, applies to a long column, the other, to a short column. Since the second version is presumably the correct version, and since in the second version the 8-letter sequence comes first, the key number 4 applies to a long column, the key number 5 to a short column in the correct version. With the foregoing deductions in mind, the solution and the reconstruction of the numerical key becomes a simple matter.

f. The text of the correct version is written out as seen in figure 25*a*. Seeing a Q in column 3 and a U in column 4, these two columns are made adjacent by sliding column 3 one interval downward, as shown in figure 25*b*. In the latter, column 7 has also been placed at the second interval to the right of column 5, because the latter yields good trigraphs with columns 3–4. Seeing the trigraph T R O near the bottom of columns 3–4–5 and the letters O and P in the same row, suggests the word TROOP. The columns are to be rearranged to make this word TROOP. There are two columns which have an O in the proper row, columns 2 and 8. The trial of combination 3–4–5–8–6, while producing TROOP in the proper row, gives bad pentagraphs in the other rows; but the combination 3–4–5–2–6 shows excellent pentagraphs, as will be seen in figure 25*c*. The words SQUADRON and HOSTILE are clearly evident; the completion of the rectangle is now a very simple matter. The result is shown in figure 25*d*. The recovery of the numerical key now will enable other cryptograms to be read directly.

1	2	3	4	5	6	7	8
		a					c
	t	o				d	o
O	H	O	N	P	R	E	O
D	D	Q	F	S	I	T	R
N	G	S	U	I	R	A	O
I	S	G	E	N	A	N	G
L	O	T	H	E	T	O	I
N	H	E	R	N	P	O	O
T	A	R	W	E	E	C	S
T	O		R		D	O	

a

1	2	3	4	5	6	7	8
							c
		a		d			o
O	T	O	N	E	R	P	O
D	H	O	F	T	I	S	R
N	D	Q	U	A	R	I	O
I	G	S	E	N	A	N	G
L	S	G	H	O	T	E	I
N	O	T	R	O	P	N	O
T	H	E	W	C	E	E	S
	A	R	R	O	D		

b

3	4	5	2	6
O	N	E	T	R
O	F	T	H	I
Q	U	A	D	R
S	E	N	G	A
G	H	O	S	T
T	R	O	O	P
E	W	C	H	E
R	R	O		D

c

3	4	7	2	6	8	1	5
O	N	E	T	R	O	O	P
O	F	T	H	I	R	D	S
Q	U	A	D	R	O	N	I
S	E	N	G	A	G	I	N
G	H	O	S	T	I	L	E
T	R	O	O	P	O	N	N
E	W	C	H	E	S	T	E
R	R	O	A	D			

d

FIGURE 25.

23. Messages with similar beginnings.—*a*. In military correspondence it is often the case that somewhat similar instructions or information must be conveyed by a superior commander to several subordinate commanders simultaneously. Such a situation frequently results in the circumstance that two or more cryptograms addressed to different stations will begin with exactly the same words. When simple columnar transposition is the system used for encipherment, then it will result, in such cases as the foregoing, that the first two or more rows of the transposition rectangle will be identical in the messages which begin alike. Therefore, the cryptograms will show identical sequences of two or more letters, distributed throughout the texts and by studying these identities the cryptanalyst is able at once not only to ascertain the width of the rectangle but also to divide up the cipher text into sections corresponding with the exact columns of the rectangle, thus eliminating the only real difficulty in solution, *viz*, the determination of which are the long columns, which the short. An example will demonstrate the short cut to solution which such a situation provides.

b. Here are two cryptograms which are assumed to have been intercepted within a few minutes of each other, the messages being addressed to two battalion commanders by the regimental commander.

CRYPTOGRAM 1

B N T S E A R K C L C E T T N B I T E R R O T A E L T N N O N N E N O
O T O K M S Z T G N Y I T D K L A N A E F T F S N P G N P A R W O I A
O F G T F C T O T D N I N O E W X E R F A S I O S T I D R R R M M A O
A R P A T O U T I O B I E O A G A A P N E I K

CRYPTOGRAM 2

B N T S E I N D O T L C E T S A F P L E R R O M O I S O E N N O N S T
I I U T O K M F E Y K P C Y I T D V S I N T A E F T F S T O N T N A R
W O A R O E E K T F C T T L T A E A N O E W X P V T I T I O S T T T F
O C M M A O O S C A N R O U T I E E L S O A G A A A B I T R T

c. The cryptanalyst, noting the similarities in the first few letters of the two messages, carefully compares the two texts, looking for additional identical sequences of letters between the cryptograms. For example, No. 1 begins with B N T S E and so does No. 2; after an interval of 4 letters in No. 1 and 5 letters in No. 2 he notes the identical sequences L C E T; after an interval of 5 letters in No. 1 and 5 letters in No. 2 he notes the identical sequences E R R O, and so on. The identities are underlined or marked in some distinctive manner throughout the texts, as shown in figure 26.

CRYPTOGRAM 1

B N T S E A R K C L C E T T N B I T E R R O T A E L T N N O N N E N O
O T O K M S Z T G N Y I T D K L A N A E F T F S N P G N P A R W O I A
O F G T F C T O T D N I N O E W X E R F A S I O S T I D R R R M M A O
A R P A T O U T I O B I E O A G A A P N E I K

CRYPTOGRAM 2

B N T S E I N D O T L C E T S A F P L E R R O M O I S O E N N O N S T
I I U T O K M F E Y K P C Y I T D V S I N T A E F T F S T O N T N A R
W O A R O E E K T F C T T L T A E A N O E W X P V T I T I O S T T T F
O C M M A O O S C A N R O U T I E E L S O A G A A A B I T R T

FIGURE 26.

d. Now it is obvious that these identities exist because the two messages begin alike, and by taking advantage of the identical portions in the cryptograms it will be possible to transcribe the texts of the latter into transposition rectangles which will not only have the identical portions in homologous positions, but also will show which are long columns, which are short. All that is necessary is to begin transcribing the texts on cross-section paper, in columns, arranging matters so that the identical sequences will fall at the tops of the columns. Thus, the first column of No. 1 will contain the letters B N T S E A R K C and the first column of No. 2 will contain the letters B N T S E I N D O T; the second column of No. 1 will contain the letters L C E T T N B I T and the second column of No. 2 will contain the letters L C E T S A F P L, and so on. It appears that the identical portion embraces the first four rows of the rectangle and runs over a number of letters on the fifth row. This is because the identical sequences consist of 4 and 5 letters. Figure 27*a* shows the identities between the first 5 columns of the two transposition rectangles. Only once in the case of this particular example does any uncertainty arise as to exactly where an identical sequence begins or ends, and that is in connection with the seventh pair of identities, involving the series of letters A E F T F S N P G N P in No. 1, and A E F T F S T O N T N in No. 2. These sequences contain 6 identical letters, but even here the uncertainty is of only a moment's duration: The initial letter A does not belong to the identical portions at the top of the transposition rectangle because the A's are needed to complete columns 6 in both rectangles. (If the A were placed at the head of column 7 in No. 1, then column 6 would lack a letter at the bottom.) Cases of "accidental identities" of course complicate the process of cutting up the text into the respective columns, but they only serve to add a small degree of interest to what would otherwise be a purely cut and dried process. The final results of the transcription into columns are shown in figure 27*b*.

1

B	L	E	N	T
N	C	R	N	O
T	E	R	O	K
S	T	O	N	M
E	T	T	N	S
A	N	A	E	Z
R	B	E	N	T
K	I	L	O	G
C	T	T	O	N

2

B	L	E	N	T
N	C	R	N	O
T	E	R	O	K
S	T	O	N	M
E	S	M	S	F
I	A	O	T	E
N	F	I	I	Y
D	P	S	I	K
O	L	O	U	P
T		E		C

FIGURE 27*a*.

e. It is obvious from a comparison of these two skeletonized matrices, and a consideration of the fact that the long columns must of necessity go to the left side, that the numbers 7 and 10 occupy the first two positions in the key, and that the numbers 2, 4, 11, and 13 occupy the last four positions in the key. By segregating and anagramming columns 7 and 10 as one group,

1

1	2	3	4	5	6	7	8	9	10	11	12	13	14
B	L	E	N	T	Y	E	A	T	N	I	M	O	A
N	C	R	N	O	I	F	R	F	O	O	M	U	G
T	E	R	O	K	T	T	W	C	E	S	A	T	A
S	T	O	N	M	D	F	O	T	W	T	O	I	A
E	T	T	N	S	K	S	I	O	X	I	A	O	P
A	N	A	E	Z	L	N	A	T	E	D	R	B	N
R	B	E	N	T	A	P	O	D	R	R	P	I	E
K	I	L	O	G	N	G	F	N	F	R	A	E	I
C	T	T	O	N	A	N	G	I	A	R	T	O	K
						P			S				

2

1	2	3	4	5	6	7	8	9	10	11	12	13	14
B	L	E	N	T	Y	E	A	T	N	I	M	O	A
N	C	R	N	O	I	F	R	F	O	O	M	U	G
T	E	R	O	K	T	T	W	C	E	S	A	T	A
S	T	O	N	M	D	F	O	T	W	T	O	I	A
E	S	M	S	F	V	S	A	T	X	T	O	E	A
I	A	O	T	E	S	T	R	L	P	T	S	E	B
N	F	I	I	Y	I	O	O	T	V	F	C	L	I
D	P	S	I	K	N	N	E	A	T	O	A	S	T
O	L	O	U	P	T	T	E	E	I	C	N	O	R
T		E		C	A	N	K	A	T		R		T

FIGURE 27*b*.

and columns 2, 4, 11, and 13 as another group, the exact positions occupied by these 6 columns are easily ascertained, as shown in figure 27*c*.

1

7	10	....	2	11	13	4
E	N	....	L	I	O	N
F	O	....	C	O	U	N
T	E	....	E	S	T	O
F	W	....	T	T	I	N
S	X	....	T	I	O	N
N	E	....	N	D	B	E
P	R	....	B	R	I	N
G	F	....	I	R	E	O
N	A	....	T	R	O	O
P	S	....				

2

7	10	....	2	11	13	4
E	N	....	L	I	O	N
F	O	....	C	O	U	N
T	E	....	E	S	T	O
F	W	....	T	T	I	N
S	X	....	S	T	E	S
T	P	....	A	T	E	T
O	V	....	F	F	L	I
N	T	....	P	O	S	I
T	I	....	L	C	O	U
N	T					

FIGURE 27c.

f. The remaining columns 1, 3, 5, 6, 8, 9, 10, 12, and 14 form a third group of columns to be anagrammed, but this is rather easy now that the columns on either side are fixed. The completed rectangles are shown in figure 27*d*.

1

7	10	3	12	6	11	4	9	5	8	2	11	13	4
E	N	E	M	Y	B	A	T	T	A	L	I	O	N
F	O	R	M	I	N	G	F	O	R	C	O	U	N
T	E	R	A	T	T	A	C	K	W	E	S	T	O
F	W	O	O	D	S	A	T	M	O	T	T	I	N
S	X	T	A	K	E	P	O	S	I	T	I	O	N
N	E	A	R	L	A	N	T	Z	A	N	D	B	E
P	R	E	P	A	R	E	D	T	O	B	R	I	N
G	F	L	A	N	K	I	N	G	F	I	R	E	O
N	A	T	T	A	C	K	I	N	G	T	R	O	O
P	S												

2

7	10	3	12	6	11	4	9	5	8	2	11	13	4
E	N	E	M	Y	B	A	T	T	A	L	I	O	N
F	O	R	M	I	N	G	F	O	R	C	O	U	N
T	E	R	A	T	T	A	C	K	W	E	S	T	O
F	W	O	O	D	S	A	T	M	O	T	T	I	N
S	X	M	O	V	E	A	T	F	A	S	T	E	S
T	P	O	S	S	I	B	L	E	R	A	T	E	T
O	V	I	C	I	N	I	T	Y	O	F	F	L	I
N	T	S	A	N	D	T	A	K	E	P	O	S	I
T	I	O	N	T	O	R	E	P	E	L	C	O	U
N	T	E	R	A	T	T	A	C	K				

FIGURE 27d.

24. **Messages with similar endings.**—*a.* What has been said at the beginning at the preceding paragraph with respect to the nature of military correspondence and the presence of identical phraseology in the messages sent by a superior commander to his subordinates also operates to produce messages in which the endings are identical. It has been noted that when two messages with similar beginnings are available for comparison, the reconstruction of the transposition rectangles and the recovery of the transposition key is an easy matter. It will now be shown that solution is an even easier matter when two messages having identical endings are available for study.

b. Given the following two cryptograms:

No. 1

```
ETRTE  EESOA  AEUNI  VAFLN  IAMND  RYHRV  MENRI
EETRO  UDCCC  OHTCY  MRREA  RHITN  DEYEN  RNERV
SRBEN  IGSKA  ILNRA  NFNAD  ALOLT  XOMAH  HRREI
```

No. 2

```
TLVSX  OPNRE  MEFDS  KYENR  UEERB  TSREH  TIANT
IVYMR  VESIR  EENEI  NOLTM  NNEDE  TROOP  UNARA
CIAAI  NSCWN  A
```

The cryptanalyst now carefully compares the two texts, searching for identical sequences of letters, but in this case instead of trying to locate identities in what may be termed a parallel progression (as in the preceding case) he searches for identical sequences of two or more letters appearing in both messages. For example, in the present case, he notes the sequence T R O forming the final trigraph of the 8th group of No. 1 and finds a similar sequence forming the initial trigraph of the 13th group of No. 2. Going through both cryptograms in this way, all the identities are marked off in some fashion, by colored crayon or underlining, as shown below. In this search for identities the cryptanalyst bears in mind that when all have been found they should be distributed at quite regular intervals throughout the text. For example, note in the following that the identities in No. 1 fall at intervals of 6 letters, with one exception; in No. 2 they fall at intervals of 4 letters, with one exception. The intervals between identities serve as a guide in finding them. After they have all been located, the identities in the cryptograms are numbered serially.

No. 1

```
E T R T E  E E S O A  A E U N I  V A F L N  I A M N D  R T H R V  M E N R I
             1                2                 3                 4
E E T R O  U D C C C  O H T C Y  M R R E A  R H I T N  D E Y E N  R N E R V
    5                   6                7                 8
S R B E N  I G S K A  I L N R A  N F N A D  A L O L T  X O M A H  H R R E I
  9                10                11                12               13
```

No. 2

```
T L V S X  O P N R E  M E F D S  K Y E N R  U E E R B  T S R E H  T I A N T
        1             2            3              4            5
I V Y M R  V E S I R  E E N E I  N O L T M  N N E D E  T R O O P  U N A R A
6            7              8            9             10             11
C I A A I  N S C W N  A
      12           13
```

c. The identities between the two cryptograms may now be equated, using for this purpose the numbers below the identities. For instance, identity 1 in cryptogram 1 matches identity 7 in cryptogram 2; identity 2 in crytogram 1 matches identity 6 in cryptogram 2, and so on. Thus:

Cryptogram 1	1	2	3	4	5	6	7	8	9	10	11	12	13
Cryptogram 2	7	6	9	2	10	5	11	3	4	12	13	1	8

d. Now cryptogram 1 has 105 letters; since the key consists of 13 numbers (indicated by the 13 identities), the rectangle for cryptogram 1 contains 12 columns of 8 letters and 1 column of 9 letters. Cryptogram 2 has 81 letters, and its rectangle contains 10 columns of 6 letters and 3 columns of 7 letters. The rectangle of cryptogram 1 has but 1 long column, whereas that of cryptogram 2 has 3 long columns. Relative to the position the last letter in each rectangle occupies in the last row of the rectangle, it is obvious that the last letter of the rectangle for cryptogram 2 is 2 letters in advance of the last letter of the rectangle for cryptogram 1. Using this difference, *viz*, 2, a cyclic sequence is generated from the series of equivalencies given above. Thus, the equivalent of identity 1 of cryptogram 1 is identity 7 of cryptogram 2, and the number 7 is placed two intervals to the right of the number 1; the equivalent of identity 7 of cryptogram 1 is identity 11 of cryptogram 2, and the number 11 is placed two intervals to the right of number 7, and so on until the following sequence is obtained:

1	2	3	4	5	6	7	8	9	10	11	12	13
1		7		11		13		8		3		9

e. The equivalent of identity 9 of cryptogram 1 is identity 4 of cryptogram 2, and the number 4 is placed between the numbers 1 and 7 in this sequence, for the sequence may be regarded as partaking of the nature of a cycle or a continuous series. From this point on, the process is the same as before, and finally the following is obtained:

1	2	3	4	5	6	7	8	9	10	11	12	13
1	4	7	2	11	6	13	5	8	10	3	12	9

f. After little experiment it becomes obvious that column 8 belongs on the extreme left because in cryptogram 1 there is only one long column, number 8, ascertained by counting the number of letters between successive identities in that message. The number 8 being at the extreme left the final actual transposition key is 8 10 3 12 9 1 4 7 2 11 6 13 5. The completely deciphered messages are shown in figure 28.

No. 1

8	10	3	12	9	1	4	7	2	11	6	13	5
H	E	A	D	R	E	D	C	O	L	U	M	N
I	N	F	A	N	T	R	Y	A	N	D	A	R
T	I	L	L	E	R	Y	M	A	R	C	H	I
N	G	N	O	R	T	H	R	E	A	C	H	E
D	S	I	L	V	E	R	R	U	N	C	R	E
E	K	A	T	S	E	V	E	N	F	O	R	T
Y	A	M	X	R	E	M	A	I	N	H	E	R
E	I	N	O	B	S	E	R	V	A	T	I	O
N												

No. 2

8	10	3	12	9	1	4	7	2	11	6	13
I	N	F	A	N	T	R	Y	P	O	I	N	T
R	E	D	C	O	L	U	M	N	P	A	S	S
E	D	S	I	L	V	E	R	R	U	N	C	R
E	E	K	A	T	S	E	V	E	N	T	W	E
N	T	Y	A	M	X	R	E	M	A	I	N	H
E	R	E	I	N	O	B	S	E	R	V	A	T
I	O	N										

FIGURE 28.

g. The possibility of the rapid solution of columnar transposition ciphers by means of the method of similar beginnings and endings, constitutes one of the most serious drawbacks to the use of transposition ciphers in military cryptography, because it is almost impossible to avoid such cases where many messages must be sent in the same key each day.

25. Solution of a single message containing a long repetition.—*a.* Sometimes a lengthy phrase or a series of numbers (spelled out in letters) is repeated within a message and if the message is enciphered by a transposition rectangle of such narrow width (in comparison with the length of the repetition) that the repeated portion forms identical sequences within the text of the cryptogram, a solution somewhat similar in principle to that explained in paragraph 24 may be achieved within a few minutes.

b. Note the following cryptogram, in which identical portions have been underlined.

CRYPTOGRAM (169 letters)

OEAEL TRS$\underset{1a}{\underline{\text{ED HN}}}$UFF $\underset{1b}{\underline{\text{RNR}}}$YF NTAED IL$\underset{2a}{\underline{\text{SMY}}}$

NCE$\underset{2b}{\underline{\text{TS LS}}}$TOC AWIAO $\underset{3a}{\underline{\text{TSLS}}}$S L$\underset{3b}{\underline{\text{EDHN}}}$ ORIIS

FEBNN $\underset{4a}{\underline{\text{UUP}}}$WE S$\underset{4b}{\underline{\text{SMY}}}$E RCNNO RVT$\underset{5a}{\underline{\text{TA O}}}GNU\underline{\text{G}}$

$\underset{5b}{\underline{\text{GTI}}}$FE RSEOM SWE$\underset{6a}{\underline{\text{RN R}}}$ASTB $\underset{6b}{\underline{\text{B OS}}}$AAA OSNOO

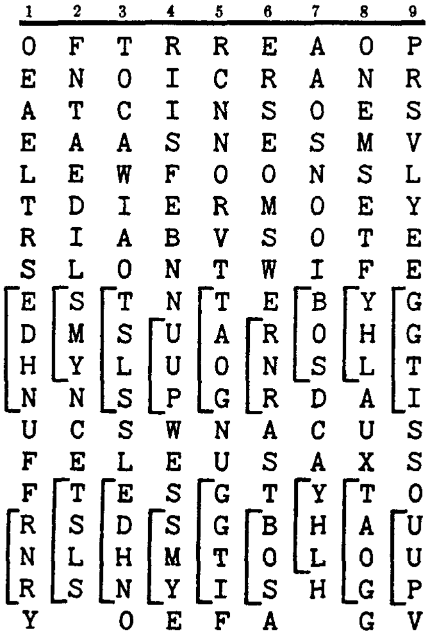
I$\underset{7a}{\underline{\text{BOS}}}$D CA$\underset{7b}{\underline{\text{YHL}}}$ HONEM SETF$\underset{8a}{\underline{\text{Y HL}}}$AUX $\underset{8b}{\underline{\text{TAO}}}$GG

PRSVL YEE$\underset{9a}{\underline{\text{GG TI}}}$SSO $\underset{9b}{\underline{\text{UUP}}}$V *c.* There are 18 segments of underlined letters, which means in this case that the rectangle is 9 columns wide, because the repeated portion in the text will give rise to two repeated sequences in each column. This means that the rectangle has 7 columns of 19 letters and 2 columns of 18 letters. The first two segments may therefore be assigned the numbers 1*a* and 1*b*, since they come from column 1; the next two segments may be assigned the numbers 2*a* and 2*b*, since they come from column 2, and so on, as shown above. Identical segments may now be equated. Thus:

1*a*	2*a*	3*a*	4*a*	5*a*	6*a*	7*a*	8*a*	9*a*
3*b*	4*b*	2*b*	9*b*	8*b*	1*b*	6*b*	7*b*	5*b*

This gives rise to the cycle 1–3–2–4–9–5–8–7–6, which is a cyclic permutation of the actual transposition key.

d. By transcribing the text into a rectangle of proper width, "cutting" the columns so as to bring the identical portions within the same rows, the result shown in figure 29 is obtained.



1	2	3	4	5	6	7	8	9
O	F	T	R	R	E	A	O	P
E	N	O	I	C	R	A	N	R
A	T	C	I	N	S	O	E	S
E	A	A	S	N	E	S	M	V
L	E	W	F	O	O	N	S	L
T	D	I	E	R	M	O	E	Y
R	I	A	B	V	S	O	T	E
S	L	O	N	T	W	I	F	E
E	S	T	N	T	E	B	Y	G
D	M	S	U	A	R	O	H	G
H	Y	L	U	O	N	S	L	T
N	N	S	P	G	R	D	A	I
U	C	S	W	N	A	C	U	S
F	E	L	E	U	S	A	X	S
F	T	E	S	G	T	Y	T	O
R	S	D	S	G	B	H	A	U
N	L	H	M	T	O	L	O	U
R	S	N	Y	I	S	H	G	P
Y		O	E	F	A		G	V

FIGURE 29.

4	6	9	1	5	3	8	2	7
R	E	P	O	R	T	O	F	A
I	R	R	E	C	O	N	N	A
I	S	S	A	N	C	E	T	O
S	E	V	E	N	A	M	A	S
F	O	L	L	O	W	S	E	N
E	M	Y	T	R	I	E	D	O
B	S	E	R	V	A	T	I	O
N	W	E	S	T	O	F	L	I
N	E	G	E	T	T	Y	S	B
U	R	G	D	A	S	H	M	O
U	N	T	H	O	L	L	Y	S
P	R	I	N	G	S	A	N	D
W	A	S	U	N	S	U	C	C
E	S	S	F	U	L	X	E	A
S	T	O	F	G	E	T	T	Y
S	B	U	R	G	D	A	S	H
M	O	U	N	T	H	O	L	L
Y	S	P	R	I	N	G	S	H
E	A	V	Y	F	O	G		

FIGURE 30.

e. Study of figure 29 shows that columns 2 and 7 are the short columns and belong on the right, either in the sequence 2–7 or 7–2. The cyclic permutation of the transposition key obtained in subparagraph *c* is 1–3–2–4–9–5–8–7–6. In order to bring the 2 and 7 adjacent in a sequence 2–7 or 7–2 one must take intervals of 5 and 4, respectively, and "decimate" the cycle, giving the following: 1–5–3–8–2–7–4–6–9 or 1–9–6–4–7–2–8–3–5. Since columns 2 and 7 belong on the right, the key must be: 4–6–9–1–5–3–8–2–7 or 8–3–5–1–9–6–4–7–2. Only a few moments are necessary to establish the correctness of the former alternative and the solution is at hand. It is as shown in figure 30.

f. A good understanding of the principles elucidated in this and the preceding paragraph will enable the student to derive for himself the procedure applicable to cases of somewhat similar nature, such as that wherein a single letter or a whole group has been omitted from the first version of a message and a second (correction message) is sent without paraphrasing the original text; or that wherein two messages are alike except for a difference in a single word (such as a number) and are cryptographed by identical transposition keys, or that wherein the numerical key has been incorrectly derived from the literal key and two versions of the same plain text are available for comparison, one based on a transposition by means of the incorrect key, the second based on a transposition by means of the correct key, both keys, however, being of the same length.

26. Solution when several cryptograms of identical length and in the same key are available.—*a.* Although the method to be described in this paragraph is included within the category of special solutions, it is of such general applicability that it might well be treated as a general solution for all transposition systems. It is based upon the very mechanics of transposition as a cryptographic scheme, *viz,* that the essential feature of the transposition method consists merely in the alterations in the positions of the elements (letters, groups of letters, or words) composing the plain text, according to a specific key. It follows, therefore, that the respective elements of two or more messages of *identical lengths,* when transposed according to the same key, will undergo identical alterations in position in the course of encipherment, and therefore all plain-text elements occupying homologous positions in the original messages will emerge in homologous positions in the cryptograms. The situation is very much like that which may be observed in the movements executed by two symmetrical groups of dancers in a chorus. Suppose each group consists of 8 dancers starting originally in definite positions relative to one another. When a movement is executed each dancer in each group performs certain evolutions; at the conclusion of the movement the 8 dancers in each group may be in quite different positions relative to one another than they were at the beginning of the movement, but the correspondingly numbered dancers in both groups find themselves in identical positions relative to their neighbors. Of course, the fact that in this analogy the groups are based upon 8's is of no significance; if the groups consisted of many more the principle would still apply. Another way of looking at the matter is to call attention to the fact that *in any type of transposition the position which a specified letter or element of the plain text will occupy in the final cryptogram is quite definitely a function of the number of letters or elements in the plain text itself.* For example, suppose that a plain-text message contains exactly 100 letters, and suppose that the transposition system and specific key is such that the 1st plain-text letter appears as the 17th cipher-text letter, the 2d plain-text letter, as the 68th, and so on; in another message of exactly 100 letters, enciphered by the same general system and specific key, it is obvious that the 1st plain-text letter must also appear as the 17th cipher-text letter, the 2d plain-text letter, as the 68th, and so on. In short, all correspondingly numbered plain-text letters in both messages will appear in identical positions in the cryptograms.

b. Granting the obvious truth of the foregoing, to what use can it be put in the solution of transposition ciphers? Simply this: *It enables the cryptanalyst to reconstruct the plain texts of cryptograms of identical length without even knowing what the transposition key or system was that produced them.* The process is not at all complicated and if there are several messages the process is very easy. It consists in superimposing the several cryptograms and *anagramming the columns* formed by the superimposition, for it is obvious that any circumstances which can be used as a guide for rearranging the letters in one of the lines of superimposed text in order to form plain text will require, and can be checked by, the results of an identical rearrangement of the corresponding letters of the other lines of superimposed text.

c. An example of the method involving the application of this general solution will now be given, using as a basis five messages assumed to have been enciphered by an unknown but complex type of transposition. It will now be shown how the security of such a system is demolished when it is used by a large number of intercommunicating commands.

d. Let the following be five cryptograms isolated from among many messages intercepted on the same day and therefore suspected of being in the same key. These five cryptograms have been isolated because they all contain exactly the same number of letters. They are here shown superimposed (fig. 31) and therefore all the letters in one column have undergone exactly the same evolutions or changes in position in the course of encipherment.

Column No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Message No. 1	I	A	A	L	N	E	O	F	S	G	T	O	G	V	E	R	A	N	O	L	N	D	U	O	D	E	I	H	I	S	A	T
Message No. 2	T	D	N	M	R	G	R	E	O	N	A	R	I	E	U	E	T	N	Y	I	T	C	O	F	E	A	I	E	U	T	T	A
Message No. 3	A	N	E	L	N	E	X	E	H	G	I	L	A	C	E	M	E	E	N	L	F	X	T	E	E	E	I	S	I	G	A	O
Message No. 4	E	E	N	E	T	S	L	N	N	F	T	C	O	I	D	O	S	E	A	I	L	F	I	G	D	W	I	A	A	R	N	O
Message No. 5	R	A	M	E	T	M	I	O	N	O	D	I	U	M	A	L	L	I	N	X	O	A	T	G	T	N	N	A	I	B	T	N

Column No.	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
Message No. 1	F	T	D	N	R	L	V	O	R	O	D	S	W	E	E	R	O	R	Q
Message No. 2	R	D	T	E	D	N	S	O	E	I	P	E	C	M	F	E	A	R	N
Message No. 3	R	W	L	L	D	L	V	V	O	R	D	E	L	O	C	H	O	T	H
Message No. 4	I	H	N	L	L	N	R	F	V	W	L	R	E	M	R	A	I	E	A
Message No. 5	H	I	T	N	I	A	S	D	R	M	S	E	C	U	I	O	V	S	A

FIGURE 31.

e. Noting a Q in message 1 column 51, the obligatory sequence Q U is assumed to be present in that message. There is in message 1 but one U, which is fortunate. Combining columns 51 and 23, the results are found to be fair (fig. 32*a*). The H T in the third row suggests a word ending in G H T, such as FIGHT, MIGHT, EIGHT, etc. Searching in message 3 for a G, two candidates are found: columns 10 and 30. The trigraphs yielded by each combination are shown in figure 32*b*. The second of the two possibilities looks much the better. The trigraph in the first row suggests the word SQUARE or SQUADRON; that in the last row suggests BATTLE or ATTALION. This means that a column with an A at the top and a T at the bottom should be sought. There is only one such column, 31. Adding it to the 30–51–23 combination gives what is shown in figure 33*a*. Looking for a column with a D at the top (for SQUAD) and either an A (for BATTALION) or an L (for BATTLE), there is only one candidate, column 22, yielding the sequences shown in figure 33*b*. Enough has been shown of the procedure to make further demonstration unnecessary. Once a good start has been made, progress is quite rapid, unless the cryptanalyst is unfortunate and arrives at a point where all the messages simultaneously terminate in complete words, without a clue as to what follows or precedes in any one of the messages. In such a contingency the only thing he can do is to try all sorts of possible continuations, either "fore" or "aft," that is, in front of the original starting point or after it, until he picks up another word which will enable him to continue. Or he may have to search for a new point of entry and build upon that, later joining this structure with the other. In the case under examination no serious difficulties are found and the entire set of five messages is reconstructed.

51	23
Q	U
N	O
H	T
A	I
A	T

FIGURE 32*a*.

10	51	23
G	Q	U
N	N	O
G	H	T
F	A	I
O	A	T

30	51	23
S	Q	U
T	N	O
G	H	T
R	A	I
B	A	T

FIGURE 32*b*.

30	51	23	31
S	Q	U	A
T	N	O	T
G	H	T	A
R	A	I	N
B	A	T	T

FIGURE 33*a*.

30	51	23	31	22
S	Q	U	A	D
T	N	O	T	C
G	H	T	A	X
R	A	I	N	F
B	A	T	T	A

FIGURE 33*b*.

f. In the course of this reconstruction the numbers applicable to the columns become assembled in proper sequence, that is, in the correct order to reproduce the plain text. This sequence, constituting the C→P sequence, is shown in figure 34 as the second row of numbers.

Term number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
C→P sequence	28	3	14	46	19	37	25	47	48	26	35	41	2	34	27	12	36	45	17	13	40	18	9	24	33	8	1	50	44	11
Message No. 1	H	A	V	E	O	R	D	E	R	E	D	R	A	T	I	O	N	W	A	G	O	N	S	O	F	F	I	R	S	T
Message No. 2	E	N	E	M	Y	D	E	F	E	A	T	E	D	D	I	R	E	C	T	I	O	N	O	F	R	E	T	R	E	A
Message No. 3	S	E	C	O	N	D	E	C	H	E	L	O	N	W	I	L	L	L	E	A	V	E	H	E	R	E	A	T	E	I
Message No. 4	A	N	I	M	A	L	D	R	A	W	N	V	E	H	I	C	L	E	S	O	F	E	N	G	I	N	E	E	R	T
Message No. 5	A	M	M	U	N	I	T	I	O	N	T	R	A	I	N	I	N	C	L	U	D	I	N	G	H	O	R	S	E	D

Term number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
C→P sequence	30	51	23	31	22	16	7	21	32	42	10	49	4	43	15	5	39	29	20	38	6
Message No. 1	S	Q	U	A	D	R	O	N	T	O	G	O	L	D	E	N	V	I	L	L	E
Message No. 2	T	N	O	T	C	E	R	T	A	I	N	A	M	P	U	R	S	U	I	N	G
Message No. 3	G	H	T	A	X	M	X	F	O	R	G	O	L	D	E	N	V	I	L	L	E
Message No. 4	R	A	I	N	F	O	L	L	O	W	F	I	E	L	D	T	R	A	I	N	S
Message No. 5	B	A	T	T	A	L	I	O	N	M	O	V	E	S	A	T	S	I	X	A	M

FIGURE 34.

g. The solution by superimposing and anagramming equal-length messages in the case of transposition constitutes a *general solution* which is applicable in all cases without exception. Indeed, the possibility of solution by this method constitutes the most serious, if not fatal, weakness of transposition as a cryptographic method, for not only is it applicable to the most complex as well as to the most simple types of transposition, but, what is much more serious, the procedure is very simple, requiring very little cryptanalytic ingenuity or expertness. The chief disadvantage of this general solution is, of course, that it is dependent upon the more or less fortuitous availability of messages of identical lengths, and while this fortunate contingency is quite frequent in a voluminous correspondence, it would naturally be better from the point of view of the cryptanalyst if this requirement were not essential in all cases. Deeper study of the subject will show that the method can still be applied in a modified way to the case of messages of almost the same lengths when the transposition is not too involved. To illustrate, a case of simple keyed-columnar transposition will be used and it will be assumed that several messages of approximately identical lengths are at hand.

h. First, take the case of two messages which have been enciphered by completely-filled rectangles, one message having, for example, one more row of letters than the other. In the discussion, the consecutive numbers 1, 2, 3, . . . will be employed as though they constituted the successive letters of a plain-text message that is being enciphered. This method of treatment is very useful in connection with studies of the mechanics of transposition ciphers in general, and especially so in the case of double transposition. Note the P→C sequences that result from the transposition:

6	2	5	7	4	1	3
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28

A

6	2	5	7	4	1	3
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21

B

P→C sequence for A.... 06 13 20 27 02 09 16 23 07 14 21 28 05 12 19 26 03 10 17 24 01 08 15 22 04 11 18 25

P→C sequence for B.... 06 13 20 02 09 16 07 14 21 05 12 19 03 10 17 01 08 15 04 11 18

FIGURE 35a

It is obvious that the two sequences may be superimposed so as to bring identical sections into superimposition. Thus:

A_____ 06 13 20 27 02 09 16 23 07 14 21 28 05 12 19 26 03 10 17 24 01 08 15 22 04 11 18 25
B_____ 06 13 20 □ 02 09 16 □ 07 14 21 □ 05 12 19 □ 03 10 17 □ 01 08 15 □ 04 11 18 □

The 7 blank spaces in the B line mark the ends of the columns in the transposition rectangle. The regularity in the distribution of the blank spaces follows from the mechanics of encipherment. If two messages were superimposed in this manner it is clear that a solution by anagramming becomes perfectly feasible. Moreover, anagramming of columns is perhaps unnecessary, for anagramming merely the letters that would occupy in line A the positions marked by the blanks in line B will yield the transposition key directly. Extension of these principles to the case in which the two rectangles differ by 2, 3, 4, . . . complete rows is obvious.

i. Taking next a case wherein two rectangles differ by one or two letters in the bottom row, it is clear that by shifting the letters of one message one or two spaces to the right (or left) from a given point will bring most of the text into proper superimposition for a solution by anagramming. Note the P→C sequences applicable to the following transpositions:

6	2	5	7	4	1	3
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	

A

6	2	5	7	4	1	3
01	02	03	04	05	06	07
08	09	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25			

B

P→C sequence for A___ 06 13 20 27 02 09 16 23 07 14 21 05 12 19 26 03 10 17 24 01 08 15 22 04 11 18 25
P→C sequence for B___ 06 13 20 02 09 16 23 07 14 21 05 12 19 03 10 17 24 01 08 15 22 04 11 18 25

FIGURE 35b

It is possible to superimpose these two sequences by shifting the sections in line B after certain numbers. Thus:

A___06 13 20 27 02 09 16 23 07 14 21 05 12 19 26 03 10 17 24 01 08 15 22 04 11 18 25
B___06 13 20 □ 02 09 16 23 07 14 21 05 12 19 □ 03 10 17 24 01 08 15 22 04 11 18 25

In the case of actual messages corresponding to the foregoing P→C sequences, superimposition of the two texts in the manner indicated would at once permit of a solution by anagramming of columns. The unknown factor, of course, is the location of the blank spaces. Where the two messages differ in length by only one or two letters brief experimentation would tell the story; where the messages differ in length by a good many letters the process would be much more difficult but not at all hopeless of fruitful results. Only a small section of text reconstructed by anagramming will soon lead to complete solution. Hence, it follows that by regulating the number of blanks to be left here and there and judicious shifting of sections of text, solution by superimposing and anagramming homologous sections of text from several messages in the same transposition key will often be possible.

j. The foregoing principles will naturally not be applicable to cases where two messages differ in length by but one letter and this small difference brings about a profound difference in the P→C sequences applicable to the messages. This is what happens often in the case of true double transposition,[1] but the principle can nevertheless be applied even here. An explanation of the procedure lies beyond the scope of the present text, however, and no more will be indicated herein concerning the matter in the case of true double transposition. However, in certain cases of combined substitution-transposition to be discussed in a later portion of this text the principles elucidated in these last few subparagraphs may be found to be applicable.

27. Reconstruction of the keys in double transposition.—*a*. Having reconstructed the plain texts of the messages solved by superimposing and anagramming, as explained in paragraph 26 *d*, *e*, can the transposition key be found? First, it is necessary to ascertain whether a single columnar transposition had been used and, if not, then the assumption will be that a double transposition had been used.

b. If a single transposition were the case, the relationship pointed out in paragraph 16*c*, concerning the existence of a constant difference between successive elements of the P→C sequence, should obtain. Having the C→P sequence, the P→C sequence may readily be established by inversion of the former. Hence, the P→C sequence is constructed by inversion, as shown in figure 36*a*.

Term number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
C→P sequence	28	3	14	46	19	37	25	47	48	26	35	41	2	34	27	12	36	45	17	13	40
	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
	18	9	24	33	8	1	50	44	11	30	51	23	31	22	16	7	21	32	42	10	49
	43	44	45	46	47	48	49	50	51												
	4	43	15	5	39	29	20	38	6												

Term number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
P→C sequence	27	13	2	43	46	51	37	26	23	41	30	16	20	3	45	36	19	22	5	49	38
	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
	35	33	24	7	10	15	1	48	31	34	39	25	14	11	17	6	50	47	21	12	40
	43	44	45	46	47	48	49	50	51												
	44	29	18	4	8	9	42	28	32												

FIGURE 36*a*

c. (1) Since there appears to be no constant difference between successive terms in the P → C sequence in figure 36*a*, single columnar transposition is ruled out and double transposition is assumed to have been employed. In passing, it is worthwhile noting that the reconstruction of the keys employed in the case of true double transposition is quite important, because it is often the case that concentrated effort directed toward the cryptanalysis of one or more messages and the subsequent recovery of the transposition keys will, of course, greatly facilitate the reading of all other messages in the same keys.

(2) There are at least four methods suited to the purpose and they will be dealt with in an order most conducive to their comprehension by the student.

(3) A preliminary to the reconstruction of the keys in the case of each of the four methods to be studied consists in establishing or ascertaining the width of either the T-1 or the T-2 matrix, usually the former, because it is easier to do.

---

[1] See Special Text No. 166, *Advanced Military Cryptography*, sec. IV.

(4) As in paragraph 26*h*, the exposition will employ matrices in which the consecutive numbers 1, 2, 3, . . . take the place of the successive plain-text letters in the T–1 matrix, because in such handling significant facts arising from the mechanics of encipherment are brought to light.

d. In order to study the effects of true double transposition on this matter of reconstructing the keys an example will be employed, involving transposition with two different keys. Let the "message" and the keys be as shown in figure 37*a:*

6	2	7	1	5	3	8	4
01	02	03	04	05	06	07	08
09	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51					

T–1

3	9	1	7	4	2	11	8	10	6	5
04	12	20	28	36	44	02	10	18	26	34
42	50	06	14	22	30	38	46	08	16	24
32	40	48	05	13	21	29	37	45	01	09
17	25	33	41	49	03	11	19	27	35	43
51	07	15	23	31	39	47				

T–2

Term number	01	02	03	04	05	06	07	08	09	10	11	12	13
P → C sequence	20	06	48	33	15	44	30	21	03	39	04	42	32
	14	15	16	17	18	19	20	21	22	23	24	25	26
	17	51	36	22	13	49	31	34	24	09	43	26	16
	27	28	29	30	31	32	33	34	35	36	37	38	39
	01	35	28	14	05	41	23	10	46	37	19	12	50
	40	41	42	43	44	45	46	47	48	49	50	51	
	40	25	07	18	08	45	27	02	38	29	11	47	

Term number	01	02	03	04	05	06	07	08	09	10	11	12	13
C → P sequence	27	47	09	11	31	02	42	44	23	34	50	38	18
	14	15	16	17	18	19	20	21	22	23	24	25	26
	30	05	26	14	43	37	01	08	17	33	22	41	25
	27	28	29	30	31	32	33	34	35	36	37	38	39
	46	29	49	07	20	13	04	21	28	16	36	48	10
	40	41	42	43	44	45	46	47	48	49	50	51	
	40	32	12	24	06	45	35	51	03	19	39	15	

FIGURE 37*a*.

Nothing in the nature of a series of constant differences between *successive* terms is now discernible in the P → C sequence. But there is, as can readily be seen, a fairly constant relationship between segments or sections of this sequence. For example, take the 1st to 6th terms of this P → C sequence (20 06 48 33 15), set them under the 29th to 34th terms (28 14 05 41 23), and find the difference between superimposed numbers. (When the minuend is less than the subtrahend the superimposed terms are disregarded.) Thus:

29th to 34th terms	28	14	05	41	23
1st to 6th terms	20	06	48	33	15
Differences	8	8		8	8

There is a constant difference between the superimposed terms. The reason for its appearance is not hard to understand if reference is made to figure 37*a* and the matter is studied in the light of the mechanics of the method of encipherment. As for the two terms 28 and the 20, while they come from different columns in the T–2 matrix, both come from the same column of the T–1 matrix, as do 14 and 06, 41 and 33, 23 and 15. But the 05 and the 48 not only come from different columns in the T–2 matrix, but also from different columns in the T–1 matrix, this representing a case where there is a transit from the bottom of one column to the top of the next column in the transposition process. Now the constant difference is in this case 8 because the superimposed terms happen to be *sequent* in the *columns* in which they fall in the T–1 matrix. If the superimposed terms are in the same column in the T–1 matrix but separated by one row, the constant difference will be 16; if separated by two rows, the constant difference will be 24, and so on. Thus, for example:

6th to 11th terms	44	30	21	03	39
29th to 34th terms	28	14	05	41	23
	16	16	16		16

Here the difference, 16, is a multiple of 8 because the superimposed terms are separated by one row in the T–1 matrix, as can be seen by referring to figure 37*a*.

e. The foregoing phenomena afford a method of ascertaining the width of the T–1 matrix in an unknown case, and, as noted above, this constitutes the first step in recovering the transposition key or keys. For if a study be made of the terms of the P → C sequence in figure 36*a*, based upon finding sections thereof which show a constant difference, the latter will correspond to either the width of the T–1 matrix or a multiple of the width. An easy way to make this study is to take a section of the P → C sequence in figure 36*a*, add 5, 6, 7, . . . (in successive steps) to each term of the selected section, and then look for repetitions between the original P → C sequence and the P → C sequence plus the additive. A beginning will be made with an assumption of a T–1 matrix of 5 columns. Since the cryptograms contain only 51 letters, all totals greater than 51 will be disregarded. Hence it is best to take a section which has a long series of low numbers, so that when the additive is applied the majority of the totals will not exceed 51. Such a series is the following (only one term in it, the 29th, is close to the maximum):

Term number	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
P → C sequence	38	35	33	24	07	10	15	01	48	31	34	39	25	14	11	17	06
P → C sequence + 5	43	40	38	29	12	15	20	06		36	39	44	30	19	16	22	11

Searching for repetitions between the P→C sequence and the P→C sequence +5, the results are negative. Trial is then made of additives 6 to 11, inclusive, with similar negative results. When an additive of 12 is applied, however, the results obtained give positive indication that the T–1 matrix is 12 columns in width. Thus:

Term number	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
P→C sequence	38	35	33	24	07	10	15	01	48	31	34	39	25	14	11	17	06
P→C sequence +12	50	47	45	36	19	22	27	13		43	46	51	37	26	23	29	18

It will be seen, on referring to figure 36*a*, that the following repetitions (with the term numbers in each of the sequences indicated) are present:

Term no. in P→C sequence + 12	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Repetitions	50	47	45	36	19	22	27	13	02	43	46	51	37	26	23	29	18
Term no. in P→C sequence	38	39	15	16	17	18	01	02	03	04	05	06	07	08	09	44	45

The width of the T–1 matrix is therefore 12 and its outlines may at once be drawn, since the total number of letters in each message, 51, indicates that there are 3 long columns of 5 letters and 9 short columns of 4 letters.

f. (1) There is another method of ascertaining the width of the T-1 matrix, which is perhaps a bit shorter and more direct than that described above. Basically both methods are the same, the one now to be presented being but another way of looking at the matter. Suppose that the differences between successive terms in the P→C sequence of figure 37*a* are calculated and set down as shown below, and then repetitions are sought in the series of differences, the latter constituting what will hereinafter be termed the *P→C interval sequence*. Thus:

Term number	01	20	03	04	05	06	07	08	09	10	11	12	13	14	15
P→C sequence	20	06	48	33	15	44	30	21	03	39	04	42	32	17	51
P→C interval sequence	−14	+42	−15	−18	+29	−14	−9	−18	+36	−35	+38	−10	−15	+34	−15

Term number	16	17	18	19	20	21	22	23	24
P→C sequence	36	22	13	49	31	34	24	09	43
P→C interval sequence	−14	−9	+36	−18	+3	−10	−15	+34	−17

Term number	25	26	27	28	29	30	31	32	33	34	35	36	37
P→C sequence	26	16	01	35	28	14	05	41	23	10	46	37	19
P→C interval sequence	−10	−15	+34	−7	−14	−9	+36	−18	−13	+36	−9	−12	−7

Term number	38	39	40	41	42	43	44	45	46	47	48	49	50	51
P→C sequence	12	50	40	25	07	18	08	45	27	02	38	29	11	47
P→C interval sequence	+38	−10	−15	−18	+11	−10	+37	−18	−25	+36	−9	−18	+36	

FIGURE 37*b*

Several repetitions are noted and underscored, in the same manner that ordinary repetitions are indicated in analogous cryptanalytic procedure. Now take the longest repetition, −14−9+36 −18, and find the terms from which it originates in the P→C sequence; a constant difference of 8 will be found. Thus:

(Term numbers 16–20)	36	22	13	49	31
(Term numbers 29–33)	28	14	05	41	23
Differences	8	8	8	8	8

The other repetitions will show the same constant difference. The terms which produce the repetitions will be found to be located in the same columns of the T-2 matrix in figure 37*a*, and reference to that figure will show that the constant difference between the sets of terms producing repetitions in the P→C interval sequence is merely the result of the mechanics of encipherment.

(2) In similar manner, if the interval sequence is constructed for the P→C sequence of figure 36*a*, the repetitions underscored in figure 36*b* are noted:

Term number	01	02	03.	04	05	06	07	08	09	10	11	12	13	14	15
P→C sequence	27	13	02	43	46	51	37	26	23	41	30	16	20	03	45
P→C interval sequence	−14	−11	+41	+3	+5	−14	−11	−3	+18	−11	−14	+4	−17	+42	−9

Term number	16	17	18	19	20	21	22	23	24
P→C sequence	36	19	22	05	49	38	35	33	24
P→C interval sequence	−17	+3	−17	+44	−11	−3	−2	−9	−17

Term number	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
P→C sequence	07	10	15	01	48	31	34	39	25	14	11	17	06	50	47
P→C interval sequence	+3	+5	−14	+47	−17	+3	+5	−14	−11	−3	+8	−11	+44	−3	−26

Term number	40	41	42	43	44	45	46	47	48	49	50	51
P→C sequence	21	12	40	44	29	18	04	08	09	42	28	32
P→C interval sequence	−9	+28	+4	−15	−11	−14	+4	+1	+33	−16	+4	

FIGURE 36*b*

Taking the sections of the P→C sequence from which the longest repetition arises and finding the constant difference between the terms involved, a width of 12 for the T–1 matrix is indicated. Thus:

(Term numbers 04–09)	43	46	51	37	26	23
(Term numbers 30–35)	31	34	39	25	14	11
Differences	12	12	12	12	12	12

This is identical with the results found by the other method. The T–1 matrix for the messages of paragraph 26*d* is therefore 12 columns in width.

g. Having ascertained the width of the T–1 matrix, the next step is to ascertain whether the width of the T–2 matrix is the same as that for the T–1, or different. If the same, the suspicion is warranted that the transposition keys for both matrices may be identical, in which case it is necessary to recover but one key. If the widths of the two matrices are different, then it is obvious that two different transposition keys are involved. Having ascertained the widths of both matrices, one can proceed to reconstruct the transposition key or keys which apply thereto. There are, as stated once before, at least four methods suitable for this purpose. They will now be taken up in turn, and each method will be explained in detail.

h. (1) In explaining the first method the discussion will be initiated with a reconsideration of figure 37*a*. If the C→P sequence established in that figure is treated as though it were plain text and enciphered by the double transposition method, using the same two transposition keys as before, an interesting phenomenon is observed. Not the following (fig. 37*c*):

6	2	7	1	5	3	8	4
27	47	09	11	31	02	42	44
23	34	50	38	18	30	05	26
14	43	37	01	08	17	33	22
41	25	46	29	49	07	20	13
04	21	28	16	36	48	10	40
32	12	24	06	45	35	51	03
19	39	15					

T–1

3	9	1	7	4	2	11	8	10	6	5
11	38	01	29	16	06	47	34	43	25	21
12	39	02	30	17	07	48	35	44	26	22
13	40	03	31	18	08	49	36	45	27	23
14	41	04	32	19	09	50	37	46	28	24
15	42	05	33	20	10	51				

T–2

FIGURE 37*c*.

Here it is seen that the numbers in the *columns* of the T–2 matrix are *consecutive*. Obviously, if the columns of this T–2 matrix were retranscribed in a matrix of the same outlines as the T–1 matrix, the numbers would be consecutive in rows and would represent the plain-text sequence 1, 2, 3, . . . , inscribed within a T–1 matrix in the normal fashion. Thus (fig. 37*d*):

6	2	7	1	5	3	8	4
01	02	03	04	05	06	07	08
09	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51					

FIGURE 37*d*.

The result of these three steps has been to effect a return to the original T–1 matrix containing plain text in normal sequence. The principal point to be noted here is the appearance of the T–2 matrix in figure 37*c*, for it is the basis of the first method for recovering the transposition keys, as well as for other operations in connection with double transposition.

(2) To demonstrate the procedure, the data afforded by figure 37*a* again will be employed. Let the C→P sequence be inscribed within a matrix of 8 columns (the width having been established by one of the methods set forth in subpars. *d–f*). Thus:

1	2	3	4	5	6	7	8
27	47	09	11	31	02	42	44
23	34	50	38	18	30	05	26
14	43	37	01	08	17	33	22
41	25	46	29	49	07	20	13
04	21	28	16	36	48	10	40
32	12	24	06	45	35	51	03
19	39	15					

FIGURE 37*e*

Find the column in which term 01 appears and set that column down *horizontally;* placing a vertical bar before and after the series of numbers to set them off as belonging to one column.

Step (1)

```
|11  38  01  29  16  06 |
```

FIGURE 37*f* (1).

Then find the column in which the term 02 appears and set it down under the row of numbers given in Step (1). Thus:

Step (2)

```
|11  38  01  29  16  06 |
        |02  30  17  07  48  35 |
```

FIGURE 37*f* (2).

Note the ascending superimposed numbers 01, 02; 29, 30; 16, 17; 06, 07. Continue to build up on this structure in the manner depicted in successive steps as follows:

Step (3)

```
            |11  38  01  29  16  06 |
                    |02  30  17  07  48  35 |
|44  26  22  13  40  03 |
```

Step (4)

```
            |11  38  01  29  16  06 |
                    |02  30  17  07  48  35 |
|44  26  22  13  40  03 |
    |27  23  14  41  04  32  19 |
```

Step (5)

```
            |11  38  01  29  16  06 |
                    |02  30  17  07  48  35 |
|44  26  22  13  40  03 |
    |27  23  14  41  04  32  19 |
                |42  05  33  20  10  51 |
```

FIGURE 37*f* (3) (4) (5).

The numbers 01 to 05, inclusive, here have formed the basis for building up the structure shown as Step (5). The next term in the sequence is 06 but it is already in the structure, as is also 07.

In the column in which 06 and 07 appear there is just room enough for 08 and 09, since the term 10 is already shown at the bottom of the column. Hence:

```
                           |11 38 01 29 16 06|
                                 |02 30 17 07 48 35|
Step (6)             |44 26 22 13 40 03|31 18 08 49 36 45|
                        |27 23 14 41 04 32 19|
                                    |42 05 33 20 10 51|

                           |11 38 01 29 16 06|
                                 |02 30 17 07 48 35|
Step (7)             |44 26 22 13 40 03|31 18 08 49 36 45|
                        |27 23 14 41 04 32 19|09 50 37 46 28 24 15|
                                    |42 05 33 20 10 51|

                           |11 38 01 29 16 06|
               |47 34 43 25 21 12 39|02 30 17 07 48 35|
Step (8)             |44 26 22 13 40 03|31 18 08 49 36 45|
                        |27 23 14 41 04 32 19|09 50 37 46 28 24 15|
                                    |42 05 33 20 10 51|
```

FIGURE 37*f* (6) (7) (8).

The process is continued in this manner until, as shown in figure 37*f*(9), all the numbers of the C→P sequence have been placed. (Here the last number is 51.)

```
                              |11 38 01 29 16 06|47 34 43 25 21 12 39|
               |47 34 43 25 21 12 39|02 30 17 07 48 35|44 26 22 13 40 03|
Step (9)             |44 26 22 13 40 03|31 18 08 49 36 45|27 23 14 41 04 32 19|
                        |27 23 14 41 04 32 19|09 50 37 46 28 24 15|
            |09 50 37 46 28 24 15|42 05 33 20 10 51|
```

FIGURE 37*f* (9).

The T–2 matrix may now be drawn within the confines of the structure shown in this last figure. The positions of vertical lines to be placed at the left and right to mark the exact outlines of the matrix may now readily be found by referring to the matrix in figure 37*e*. It is obvious that the column with the terms 11–15 belongs at the extreme left of the T–2 matrix, the column with the terms 21–24 belongs at the extreme right. The transposition key for the matrix may then be established directly from the matrix itself, by following the sequence of numbers in the columns. Thus:

```
                    3  9  1  7  4  2 11  8 10  6  5
                   11 38 01 29 16 06|47 34 43 25 21 12 39|
   |47 34 43 25 21 12 39|02 30 17 07 48 35|44 26 22 13 40 03|
         |44 26 22 13 40 03|31 18 08 49 36 45|27 23 14 41 04 32 19|
            |27 23 14 41 04 32 19|09 50 37 46 28 24 15|
|09 50 37 46 28 24 15|42 05 33 20 10 51
```

FIGURE 37*g*.

Reference to figure 37*c* will show the exact correspondence between the T–2 transposition key and the T–2 matrix of figure 37*g* with these same elements indicated in figure 37*c*. The transposition key for the T–1 matrix in figure 37*e* can now easily be derived from figure 37*g*. It must be as follows:

6	2	7	1	5	3	8	4
27	47	09	11	31	02	42	44
23	34	50	38	18	30	05	26
14	43	37	01	08	17	33	22
41	25	46	29	49	07	20	13
04	21	28	16	36	48	10	40
32	12	24	06	45	35	51	03
19	39	15					

FIGURE 37*h*.

This transposition key and T–1 matrix are identical with the key and T–1 matrix of figure 37*c*.

i. Note the application of the foregoing method to the C→P sequence shown in figure 36*a* in connection with the messages solved in paragraph 26*d–e*. It has already been found that the width of the T–1 matrix is 12. The C→P sequence of figure 36*a* is therefore inscribed within a matrix of 12 columns.

1	2	3	4	5	6	7	8	9	10	11	12
28	03	14	46	19	37	25	47	48	26	35	41
02	34	27	12	36	45	17	13	40	18	09	24
33	08	01	50	44	11	30	51	23	31	22	16
07	21	32	42	10	49	04	43	15	05	39	29
20	38	06									

FIGURE 36*c*.

The process explained in subparagraph *h* (2) above is now applied. The successive steps have been omitted but the final result is shown herewith:

```
               | 14   27   01   32   06 | 19   36   44   10 |
| 48   40   23   15 | 28   02   33   07   20 | 37   45   11   49 |
     | 41   24   16   29 | 03   34   08   21   38 | 46   12   50   42 |
          | 25   17   30   04 | 35   09   22   39 | 47   13   51   43 |
          | 26   18   31   05 |
```

FIGURE 36*d*.

All the numbers from 01 to 51, inclusive, are contained within this structure. Extending it to the left or right to make the T–2 matrix complete, by referring to the T–1 matrix, it is found that the structure must be made as shown herewith:

```
                  4     7     1     8     2     5     9    11     3    12    10     6
                 14   27   01   32   06 | 19   36   44   10 | 48   40   23   15 |
| 48   40   23   15 | 28   02   33   07   20 | 37   45   11   49 | 41   24   16   29 |
     | 41   24   16   29 | 03   34   08   21   38 | 46   12   50   42 | 25   17   30   04 |
          | 25   17   30   04 | 35   09   22   39 | 47   13   51   43 | 26   18   31   05 |
          | 26   18   31   05
```

FIGURE 36*e*.

The transposition key for the T–1 matrix is now found to be as indicated at the top of figure 36*e*.

j. A second method for reconstructing the keys will now be explained. To demonstrate this method the data afforded by figure 37*a* will again be employed. Going back to the point where the P→C interval sequence for this example was established in subparagraph *f*(1) above, the terms, in figure 37*b*, which gave rise to the thrice-appearing repetition in the interval sequence (−10 −15 +34) are found to be as follows:

1st appearance (term numbers 12–16)	42	32	17	51
2d appearance (term numbers 21–25)	34	24	09	43
3d appearance (term numbers 25–29)	26	16	01	35

FIGURE 37*i*.

These sequences may be rearranged so as to bring the numbers in ascending order within columns. Thus:

26	16	01	35
34	24	09	43
42	32	17	51

FIGURE 37*j*.

The constant difference, 8, within the columns of this structure is, of course, the same constant difference as was found before, and corresponds with the width of the T–1 matrix. It derives from the T–1 matrix, as may be seen on referring to figure 37*a*. The columns of the structure in figure 37*j* are seen to be portions of the T–1 matrix, lying in the following positions in that matrix:

01	.	.	.	.	.	.	.	.	.	.
09	.	.	.	.	.	.	.	.	.	16
17	.	.	.	.	.	.	.	.	.	24
.	26	.	.	.	.	.	.	.	.	32
.	34	35	.	.	.	.	.	.	.	.
.	42	43	.	.	.	.	.	.	.	.
.	.	51								

FIGURE 37*k*.

In the T–2 matrix these numbers fall in the following positions:

.	.	.	.	.	.	.	.	.	26	34
42	.	.	.	.	.	.	.	.	16	24
32	.	.	.	.	.	.	.	.	01	09
17	.	.	.	.	.	.	.	.	35	43
51	.	.	.	.	.	.				

FIGURE 37*l*.

Now if the dimensions of the T–2 matrix were *unknown*, these numbers could nevertheless be placed in a skeletonized T–2 matrix as follows:

26	34	42
16	24	32
01	09	17
35	43	51

FIGURE 37*m*(1).

and the block of numbers could be extended on both sides by referring to the T–1 matrix in figure 37*a*. Thus:

```
       | 02  10  18 | 26  34  42 | 50
               | 08 | 16  24  32 | 40  48 |
                    | 01  09  17 | 25  33  41  49 |
  | 03  11  19  27 | 35  43  51 |
```

FIGURE 37*m*(2).

This structure may next be extended by referring to the P→C sequence in figure 37*a*, by completing the partial columns of the structure.

```
 44 | 02  10  18 | 26  34  42 | 50  06  14  22
 30   38  46 | 08 | 16  24  32 | 40  48 | 05  13
 21   29  37  45 | 01  09  17 | 25  33  41  49
|03   11  19  27 | 35  43  51 | 07  15  23  31
```

FIGURE 37*m*(3).

Again the structure may be extended by referring to the T–1 matrix. Thus:

```
04  12  20  28  36  44  02  10  18 | 26  34  42 | 50  06  14  22  30  38  46
        06  14  22  30  38  46  08 | 16  24  32 | 40  48  05  13  21  29  37  45
            05  13  21  29  37  45 | 01  09  17 | 25  33  41  49
                    03  11  19  27 | 35  43  51 | 07  15  23  31  39  47
```

FIGURE 37*m*(4).

Noting the appearance of the term 06 in the 1st row and also in the 2d row of the structure, the latter may be transcribed as follows:

```
+--------------------------------------------+
| 04  12  20  28  36  44  02  10  18  26  34 | 42  50  06  14  22  30  38  46
| 42  50  06  14  22  30  38  46  08  16  24 | 32  40  48  05  13  21  29  37  45
| 32  40  48  05  13  21  29  37  45  01  09 | 17  25  33  41  49
| 17  25  33  41  49  03  11  19  27  35  43 | 51  07  15  23  31  39  47
| 51  07  15  23  31  39  47 +---------------+
+----------------------------+
```

FIGURE 37*m* (5).

By referring to the T–1 matrix of figure 37*a* and the foregoing structure, the key for T–1 can be recovered. It is 6–2–7–1–5–3–8–4. By referring to the P→C sequence in figure 37*a* the key for the T–2 matrix just constructed may be established. It is 3–9–1–7–4–2–11–8–10–6–5.

k. Applying this method to the messages solved in paragraph 26*d–e*, the steps are as follows: The width of the T–1 matrix has been established as being 12. The P→C interval sequence in figure 36*b* shows the repetition +3 +5 —14 —11 —3 appearing two times; the repetition +3 +5 —14 appearing three times; and the repetition —14 —11 appearing three times. The terms giving rise to these repetitions are arranged in a structure with ascending numbers within the columns. Thus:

```
07  10  15  01
        27  13  02
31  34  39  25  14  11
43  46  51  37  26  23
```

FIGURE 36*f*.

The constant difference, 12, indicates a T–1 matrix of 12 columns. The matrix is prepared:

1	2	3	4	5	6	7	8	9	10	11	12
01	02	03	04	05	06	07	08	09	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48
49	50	51									

FIGURE 36*g*.

The terms within the columns of the structure in figure 36*f* are transcribed into rows (of the skeletonized T–2 matrix):

07		31	43
10		34	46
15	27	39	51
01	13	25	37
	02	14	26
		11	23

FIGURE 36*h*(1).

This structure is extended by referring to the T–1 matrix (figure 36*g*):

	1	2	3	4	5	6	7
1		07	19	31	43		
2		10	22	34	46		
3	03	15	27	39	51		
4		01	13	25	37	49	
5			02	14	26	38	50
6				11	23	35	47

FIGURE 36*h*(2).

Noting that the initial terms of the P→C sequence in figure 36*a* (27 13 02) are present in this structure (in the 3d column) this gives the top of the T–2 matrix as coinciding with the 3d row of the structure. The P→C sequence in figure 36*a* reads 27 13 02 43 46 . . .; the 43 and 46 are also in the structure in figure 36*h* (2) in the 1st and 2d rows, column 5; hence the structure in figure 36*h*(2) can be rearranged thus:

	03	15	27	39	51		
		01	13	25	37	49	
			02	14	26	38	50
07	19	31	43	11	23	35	47
10	22	34	46				

FIGURE 36*h*(3).

The structure may now be extended by referring to the P→C sequence in figure 36*a*:

35	03	15	27	39	51	05	17
33	45	01	13	25	37	49	06
24	36	48	02	14	26	38	50
07	19	31	43	11	23	35	47
10	22	34	46				

FIGURE 36*h*(4).

Thus, by referring alternately to the P→C sequence and the T–1 matrix the structure is extended to the following:

```
35 | 03  15  27  39  51 | 05  17  29  41 | 09  21  33 | 45
33 | 45 | 01  13  25  37  49 | 06  18  30  42 | 12  24 | 36  48
24 | 36  48 | 02  14  26  38  50 | 04  16  28  40 | 07 | 19  31  43 |
07 | 19  31  43 | 11  23  35  47 | 08  20  32  44 | 10 | 22  34  46 |
10 | 22  34  46
```

FIGURE 36*h*(5).

It will be noted that the first number to the right of each vertical bar is one of the numbers from 1 to 12, indicating that all the columns of the T–1 matrix are now represented in the T–2 structure. It is now easy to write the transposition key over the T–1 matrix: 4–7–1–8–2–5–9–11–3–12–10–6. By following the numbers in the P→C sequence the transposition key for the T–2 matrix is given directly; it is the same as for the T–1 matrix.

l. (1) A third method for reconstructing the transposition keys will now be set forth. It will first be explained in connection with the artificial example in figure 37*a*. It has been noted how the width of the T–1 matrix can be ascertained from a study of the P→C sequence; the work in connection with figure 37*a* and subparagraph *e* give an indicated width of 8 for the T–1 matrix in this case.

(2) Let the additive 8 (found in subpars. *d* and *f*) be applied to the entire P→C sequence of figure 37*a*, and then let the identities between the two sequences be underscored and numbered, as shown in figure 37*n*:

```
(A) P→C sequence..............{ 20  06  48  33  15 |44  30  21  03  39 |04 |42  32  17  51 |
                                        1                   2                    3
                                36  22  13  49  31 |34  24  09 | 43 | 26  16  01  35 |
                                        4              5               6
                                28  14  05  41  23 |10  46  37  19 | 12 |50  40  25  07 |
                                        7                8                  9
                                18  08  45  27 | 02  38  29  11  47 |
                                      10                11

(B) P→C sequence+8............{ 28  14  56  41  23 |52  38  29  11  47 |12 |50  40  25  59 |
                                        1                   2                    3
                                44  30  21  57  39 |42  32  17  51 | 34  24  09 | 47
                                        4                5              6
                                36  22  13  49  31 |18  54  45  27 | 20  58  48  33  15 |
                                        7                8                  9
                                26  16  53  35 |10  46  37  19  55 |
                                      10                11
```

FIGURE 37*n*.

If now the procedure explained in paragraph 16*k*, 24*c* to *f*, and 25*c* to *e* is applied to the repetitions noted in figure 37*n*, it becomes clear that the T–2 matrix in this case must have 11 columns. The transposition key for that matrix is then established, as follows: [2]

B	1	2	3	4	5	6	7	8	9	10	11
A	7	11	9	2	3	5	4	10	1	6	8
Chain	1	7	4	2	11	8	10	6	5	3	9

---

[2] It is to be noted that the B sequence (that is, the P→C sequence plus the additive) must be used as the base, otherwise the chain of equivalents will be a *reversal* of the correct chain.

This is a cyclic permutation of the key for the T-2 matrix; to obtain the actual key it is necessary merely to fix the position of one of the key numbers with respect to the matrix. It is easy to find which number belongs at the extreme left or extreme right of the matrix Only a few minutes experimentation with the key and the T-2 matrix gives the correct starting point for the key, which is found to be 3-9-1-7-4-2-11-8-10-6-5.

(3) The recovery of the transposition key for the T-1 matrix is now a simple matter. Its width having been established as 8 columns, a mere transcription of the P→C sequence numbers from the T-2 matrix into the T-1 matrix gives the key 6-2-7-1-5-3-8-4. The two keys and matrices are found to be different.

(4) The procedure set forth in this subparagraph is applicable without modification to the case where the two transposition matrices are the same and have the same transposition key. This will be noted in the following demonstration of the recovery of the matrices and keys for the messages solved in paragraph 26*d* and *e* by anagramming. It has already been shown how the width of the T-1 matrix was ascertained as being 12 columns (subpar. *f*). The additive 12 is then applied to the entire P→C sequence, identities are established between sections of the original sequence and sections of the sequence + 12, and these identical sections are equated in the usual manner, leading to the establishment of a cyclic permutation of the transposition key for the T-2 matrix. Thus (fig. 36*i*):

	A. (P→C sequence)	B. (P→C sequence + 12)
1	27 13 02 43 46	39 25 14 55 58
2	51 37 26 23	63 49 38 35
3	41 30 16 20	53 42 28 32
4	03 45 36 19 22	15 57 48 31 34
5	05 49 38 35	17 61 50 47
6	33 24 07 10	45 36 19 22
7	15 01 48 31 34	27 13 60 43 46
8	39 25 14 11	51 37 26 23
9	17 06 50 47	29 18 62 59
10	21 12 40 44	33 24 52 56
11	29 18 04 08	41 30 16 20
12	09 42 28 32	21 54 40 44

FIGURE 36*i*.

B	1	2	3	4	5	6	7	8	9	10	11	12
A	8	5	12	7	9	4	1	2	11	6	3	10
Chain	1	8	2	5	9	11	3	12	10	6	4	7

Since sections 1, 4, and 7 of the P→C sequence contain 5 terms (=long columns), the other sections only 4 (=short columns), it follows that the key numbers 4-7-1 go to the left and the actual key for the T-2 matrix is 4-7-1-8-2-5-9-11-3-12-10-6. Since the number of elements in this key is the same as in the key for the T-1 matrix, it is likely that the same key is employed for both transpositions. Simple experiment will quickly verify this assumption and the transposition matrices for the first of the 5 messages of paragraph 26*d* may be seen in the following (figure 36*j*):

4	7	1	8	2	5	9	11	3	12	10	6
1 H	2 A	3 V	4 E	5 O	6 R	7 D	8 E	9 R	10 E	11 D	12 R
13 A	14 T	15 I	16 O	17 N	18 W	19 A	20 G	21 O	22 N	23 S	24 O
25 F	26 F	27 I	28 R	29 S	30 T	31 S	32 Q	33 U	34 A	35 D	36 R
37 O	38 N	39 T	40 O	41 G	42 O	43 L	44 D	45 E	46 N	47 V	48 I
49 L	50 L	51 E									

T-1

T-2

4	7	1	8	2	5	9	11	3	12	10	6
3 V	15 I	27 I	39 T	51 E	5 O	17 N	29 S	41 G	9 R	21 O	33 U
45 E	1 H	13 A	25 F	37 O	49 L	6 R	18 W	30 T	42 O	12 R	24 O
36 R	48 I	2 A	14 T	26 F	38 N	50 L	4 E	16 O	28 R	40 O	7 D
19 A	31 S	43 L	11 D	23 S	35 D	47 V	8 E	20 G	32 Q	44 D	10 E
22 N	34 A	46 N									

Cryptogram.................. I A A L N E O F S G etc.
P→C sequence............ 27 13 2 43 46 51 37 26 23 41 etc.

FIGURE 36*j*.

m. (1) A fourth and possibly the most elegant solution to the problem of reconstructing the keys for double transposition will now be presented.[3] Reference will be made to the two matrices and keys shown in figure 36*j*. Let the P→C$_1$ and P→C$_2$ sequences resulting from the first and the second transpositions, respectively, be shown, as seen below:

1. Term number.......	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
2. P→C$_1$ sequence....	03	15	27	39	51	05	17	29	41	09	21	33	45	01	13	25	37
P→C$_2$ sequence....	27	13	02	43	46	51	37	26	23	41	30	16	20	03	45	36	19
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
	49	06	18	30	42	12	24	36	48	02	14	26	38	50	04	16	28
	22	05	49	38	35	33	24	07	10	15	01	48	31	34	39	25	14
	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
	40	07	19	31	43	11	23	35	47	08	20	32	44	10	22	34	46
	11	17	06	50	47	21	12	40	44	29	18	04	08	09	42	28	32

FIGURE 36*k*

---

[3] The basic principles underlying this fourth and most important method were discovered and first presented 1934 by Solomon Kullback, Ph. D., then Junior Cryptanalyst, S. I. S.

A casual examination of these three rows of numbers discloses an interesting *invariant* relationship between any pair of superimposed numbers in rows 1 and 2 and in rows 2 and 3. For instance, take the very first pair, $\genfrac{}{}{0pt}{}{01}{03}$ in rows 1 and 2; in rows 2 and 3 the same pair of superimposed numbers will be found (under term No. 14). This same relationship exists between all the superimposed pairs in rows 1–2 and 2–3.

(2) Given only the third row of numbers in figure 36*k*, that is, the P→C$_2$ sequence (which has heretofore been designated merely as the P→C sequence), obtained as a result of a solution by superimposing and anagramming several messages, it is not difficult to reconstruct the second row, the P→C$_1$ sequence. The width of the T–1 matrix can be ascertained by either of the two methods indicated in subparagraphs *e* and *f*. It is now known to be 12. A 12-column matrix is therefore constructed, containing 51 cells numbered in the normal manner. This will, of course, give the T–1 matrix seen in figure 36*j*, but without the transposition key or the letters in the cells. Thus:

1	2	3	4	5	6	7	8	9	10	11	12
01	02	03	04	05	06	07	08	09	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48
49	50	51									

FIGURE 36*l*

The invariant relationship pointed out in subparagraph (1) above may now be used to establish the T–1 key. Since the key is known to contain 12 elements, a start may be made with any one of 12 possibilities. Suppose that the key begins with 1. The first five terms in the P→C$_1$ sequence would be as indicated herewith:

1. Term number	01	02	03	04	05
2. P→C$_1$ sequence	01	13	25	37	49
3. P→C$_2$ sequence	27	13	02	43	46

Two "conflicts" or contradictions are at once manifested: $\genfrac{}{}{0pt}{}{01}{01}$ in rows 1 and 2, $\genfrac{}{}{0pt}{}{01}{27}$ in rows 2 and 3; also, $\genfrac{}{}{0pt}{}{02}{13}$ in rows 1 and 2, $\genfrac{}{}{0pt}{}{13}{13}$ in rows 2 and 3. The conclusion is obvious that the key number 1 does not occupy the 1st position in the transposition key. Suppose key number 1 belongs in the 2d position in the key. The superimposed sequences are then as follows:

1. Term number	01	02	03	04	05
2. P→C$_1$ sequence	02	14	26	38	50
3. P→C$_2$ sequence	27	13	02	43	46

Here again two conflicts are noted: $\genfrac{}{}{0pt}{}{01}{02}$ in rows 1 and 2, $\genfrac{}{}{0pt}{}{26}{02}$ in rows 2 and 3; $\genfrac{}{}{0pt}{}{02}{14}$ in rows 1 and 2, $\genfrac{}{}{0pt}{}{02}{27}$ in rows 2 and 3. Only a single contradiction is sufficient to permit of discarding an hypothesis. The key number 1 does not occupy the 2d position in the key. A trial is made of the 3d position for key number 1. The results are as follows:

1. Term number	01	02	03	04	05
2. P→C$_1$ sequence	03	15	27	39	51
3. P→C$_2$ sequence	27	13	02	43	46

Here there are no contradictions and one check or corroboration: $\genfrac{}{}{0pt}{}{03}{27}$ in rows 1 and 2, $\genfrac{}{}{0pt}{}{03}{27}$ in rows 2 and 3. If key number 1 really occupies the 3d position in the key, then the superimposition data given in the last set of rows of superimposed numbers may be employed, by transferring the data to the proper positions in the skeletonized figure 36*m* (1):

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
2. $P\rightarrow C_1$ sequence	03	15	27	39	51	05								01			
3. $P\rightarrow C_2$ sequence	27	13	02	43	46	51	37	26	23	41	30	16	20	03	45	36	19

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1. Term number	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
2. $P\rightarrow C_1$ sequence										02					04		
3. $P\rightarrow C_2$ sequence	22	05	49	38	35	33	24	07	10	15	01	48	31	34	39	25	14

	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1. Term number	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
2. $P\rightarrow C_1$ sequence																	
3. $P\rightarrow C_2$ sequence	11	17	06	50	47	21	12	40	44	29	18	04	08	09	42	28	32

FIGURE 36*m* (1)

It then becomes at once possible, by referring to the T-1 matrix, to insert more numbers in the $P\rightarrow C_1$ sequence. Thus:

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
2. $P\rightarrow C_1$ sequence	03	15	27	39	51	05	17	29	41					01	13	25	37
3. $P\rightarrow C_2$ sequence	27	13	02	43	46	51	37	26	23	41	30	16	20	03	45	36	19

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1. Term number	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
2. $P\rightarrow C_1$ sequence	49									02	14	26	38	50	04	16	28
3. $P\rightarrow C_2$ sequence	22	05	49	38	35	33	24	07	10	15	01	48	31	34	39	25	14

	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1. Term number	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
2. $P\rightarrow C_1$ sequence	40																
3. $P\rightarrow C_2$ sequence	11	17	06	50	47	21	12	40	44	29	18	04	08	09	42	28	32

FIGURE 36*m* (2)

The new placements now permit of placing numbers in the $P\rightarrow C_1$ sequence. For example, $\genfrac{}{}{0pt}{}{07}{17}$ in rows 1 and 2 permit of placing the number 07 above the number 17 in the $P\rightarrow C_2$ sequence; $\genfrac{}{}{0pt}{}{08}{29}$ in rows 1 and 2 permit of placing the number 08 above the number 29 in the $P\rightarrow C_2$ sequence, and so on. In only a few moments the entire $P\rightarrow C_1$ sequence can be established. Thus:

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
2. $P\rightarrow C_1$ sequence	03	15	27	39	51	05	17	29	41	09	21	33	45	01	13	25	37
3. $P\rightarrow C_2$ sequence	27	13	02	43	46	51	37	26	23	41	30	16	20	03	45	36	19

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1. Term number	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
2. $P\rightarrow C_1$ sequence	49	06	18	30	42	12	24	36	48	02	14	26	38	50	04	16	28
3. $P\rightarrow C_2$ sequence	22	05	49	38	35	33	24	07	10	15	01	48	31	34	39	25	14

	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1. Term number	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
2. $P\rightarrow C_1$ sequence	40	07	19	31	43	11	23	35	47	08	20	32	44	10	22	34	46
3. $P\rightarrow C_2$ sequence	11	17	06	50	47	21	12	40	44	29	18	04	08	09	42	28	32

FIGURE 36*n*.

(3) The determination of the T-1 key is now a very simple matter. Since it is known that the key has 12 numbers, it is only necessary to note in the $P\rightarrow C_1$ sequence the relative order of the numbers 1 to 12. It is as follows:

1	2	3	4	5	6	7	8	9	10	11	12
3	5	9	1	6	12	2	4	7	11	8	10

This is merely the inverse of the actual key; the latter may be obtained by inversion. Thus:

1	2	3	4	5	6	7	8	9	10	11	12
4	7	1	8	2	5	9	11	3	12	10	6

Comparison of this key with the T–1 key shown in figure 36*j* will establish the identity of the two. The determination of the T–2 key is obvious, having the T–1 at hand. In this case both matrices and keys are identical.

n. Attention will be directed to a further interesting phenomenon in this case. Referring to figure 36*n*, if chains of equivalents are constructed between elements of the 1st and 3d rows only, the following two chains are obtained:

01 27 15 45 18 22 35 11 30 31 34 14 03 02 13 20 49 42 40 21 38 50 28
04 43 44 29 48 09 23 33 25 07 37 06 51 32 39 47 08 26 10 41 12 16 36 17 19 05 46

FIGURE 36*o*.

All the terms of the $P \rightarrow C_2$ sequence are represented, except the number 24, which stands by itself. If now each of these chains is slid against itself, when properly juxtaposed, the superimposed pairs are identical with those in rows 1 and 2 in figure 36*n*. Note the following:

(1) 01 27 15 45 18 22 35 11 30 31 34 14 03 02 13 20 49 42 40 21 38 50 28
03 02 13 20 49 42 40 21 38 50 28 01 27 15 45 18 22 35 11 30 31 34 14

(2) 04 43 44 29 48 09 23 33 25 07 37 06 51 32 39 47 08 26 10 41 12 16 36 17 19 05 46
39 47 08 26 10 41 12 16 36 17 19 05 46 04 43 44 29 48 09 23 33 25 07 37 06 51 32

FIGURE 36*p*.

The application of the foregoing phenomena in the case under study is obvious. Here it is not even necessary to ascertain the width of the T–1 matrix before proceeding to try to establish the T–1 key. Of course, the number of chains which may be established will vary with the

6	2	7	1	5	3	8	4
01	02	03	04	05	06	07	08
09	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51					

A (=T–1)

3	9	1	7	4	2	11	8	10	6	5
01	02	03	04	05	06	07	08	09	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41	42	43	44
45	46	47	48	49	50	51				

B

3	9	1	7	4	2	11	8	10	6	5
04	12	20	28	36	44	02	10	18	26	34
42	50	06	14	22	30	38	46	08	16	24
32	40	48	05	13	21	29	37	45	01	09
17	25	33	41	49	03	11	19	27	35	43
51	07	15	23	31	39	47				

C (=T–2)

FIGURE 37*o*.

specific matrices and keys, but the general principles herein presented may nevertheless be applied. In some cases it may be necessary to juxtapose two different chains obtained by equating terms from rows 1 and 3, rather than juxtaposing one chain against itself. Only a few minutes experimentation will be necessary to establish contradictions which will permit of discarding fallacious hypotheses.

o. (1) In the foregoing explanation, the two transposition keys and matrices were identical. Even when they are different the same principles, with minor modifications, may be applied. The matrices and keys of figure 37*a* will again be employed to demonstrate the necessary modifications.

(2) First, prepare the two matrices with consecutive numbers in the cells of *both* matrices, as shown at A and B in figure 37*o* and then prepare the T-2 matrix, shown at C.

(3) Write the $P \rightarrow C_1$ sequence for T-1, under it write the $P \rightarrow C_1$ for T-2, and under the latter write the $P \rightarrow C_2$ sequence for the final cryptogram. Thus:

1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14
2. $P \rightarrow C_1$ sequence for T-1	04	12	20	28	36	44	02	10	18	26	34	42	50	06
3. $P \rightarrow C_1$ sequence for T-2	03	14	25	36	47	06	17	28	39	50	01	12	23	34
4. $P \rightarrow C_2$ sequence	20	06	48	33	15	44	30	21	03	39	04	42	32	17

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	14	22	30	38	46	08	16	24	32	40	48	05	13	21
	45	05	16	27	38	49	11	22	33	44	10	21	32	43
	51	36	22	13	49	31	34	24	09	43	26	16	01	35

	29	30	31	32	33	34	35	36	37	38	39	40	41	42
	29	37	45	01	09	17	25	33	41	49	03	11	19	27
	04	15	26	37	48	08	19	30	41	02	13	24	35	46
	28	14	05	41	23	10	46	37	19	12	50	40	25	07

	43	44	45	46	47	48	49	50	51
	35	43	51	07	15	23	31	39	47
	09	20	31	42	07	18	29	40	51
	18	08	45	27	02	38	29	11	47

FIGURE 37*p*.

Note, now, the invariant relationship between rows 1-2 and 3-4. The same phenomenon is here manifested as was encountered in the preceding case where the T-1 and T-2 matrices and keys were identical. It follows, therefore, that the principles elucidated under subparagraph *m* may be applied, with some modifications, also to the case where different keys and matrices are employed for double transposition. The width of the T-1 matrix may be ascertained in the manner already indicated; an assumption is made as to the position occupied by key number 1 of the T-1 key; this assumption provides data for making an assumption as to the width of the T-2 matrix. When the correct pair of assumptions is made, the data in rows 1 and 2 are corroborated by those in rows 3 and 4. From that point on the rest is easy and follows along the same lines as before.

p. (1) The procedure will be illustrated by employing the P→C sequence in figure 37*b* (which is the same as that labelled $P \rightarrow C_2$ sequence in figure 37*p*), it being assumed that nothing is known about the matrices, and that the sequence was obtained from a solution by superimposing and anagramming several messages of identical length.

(2) The width of the T-1 matrix is established as 8 and the T-1 matrix set down:

1	2	3	4	5	6	7	8
01	02	03	04	05	06	07	08
09	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51					

T-1

FIGURE 37q.

(3) Assuming that key number 1 occupies the first position in the T-1 key, the numbers are inserted in row 2, representing the beginning of the P→C₁ sequence for T-1. The superimposed pairs in rows 1 and 2 are distributed in rows 3 and 4, with the results shown in figure 37r(1).

1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14
2. P→C₁ sequence for T-1	01	09	17	25	33	41	49							
3. P→C₁ sequence for T-2				05										03
4. P→C₁ sequence	20	06	48	33	15	44	30	21	03	39	04	42	32	17

1. Term number	15	16	17	18	19	20	21	22	23	24	25	26	27	28
2. P→C₁ sequence for T-1														
3. P→C₁ sequence for T-2					07				02				01	
4. P→C₁ sequence	51	36	22	13	49	31	34	24	09	43	26	16	01	35

1. Term number	29	30	31	32	33	34	35	36	37	38	39	40	41	42
2. P→C₁ sequence for T-1														
3. P→C₁ sequence for T-2				06									04	
4. P→C₁ sequence	28	14	05	41	23	10	46	37	19	12	50	40	25	07

1. Term number	43	44	45	46	47	48	49	50	51
2. P→C₁ sequence for T-1									
3. P→C₁ sequence for T-2									
4. P→C₁ sequence	18	08	45	27	02	38	29	11	47

FIGURE 37r (1).

(4) An attempt is now made to construct a T-2 matrix which will produce the distribution and spacing of the numbers in row 3. For example, from the position of the number 05 the matrix would have to be of such dimensions that there are short columns of 2 letters and long columns of 3 letters; or short columns of 3 letters and long columns of 4 letters. The former hypothesis can be discarded at once, for the intervals between the numbers 03, 07, 02, 01, and 06 in row 3 make it untenable. The latter hypothesis may also be discarded, for the intervals between 03 and 07 and between 01 and 06 make it impossible. Hence key number 1 cannot occupy the first position in the T-1 key. Position 2 is assumed for key number 1 and the procedure repeated, also without good results. Note what happens when position 4 is assumed for key number 1 in the T-1 key:

1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14
2. P→C_1 sequence for T-1	04	12	20	28	36	44								
3. P→C_2 sequence for T-2	03					06					01			
4. P→C_2 sequence	20	06	48	33	15	44	30	21	03	39	04	42	32	17
1. Term number	15	16	17	18	19	20	21	22	23	24	25	26	27	28
2. P→C_1 sequence for T-1														
3. P→C_2 sequence for T-2		05												
4. P→C_2 sequence	51	36	22	13	49	31	34	24	09	43	26	16	01	35
1. Term number	29	30	31	32	33	34	35	36	37	38	39	40	41	42
2. P→C_1 sequence for T-1														
3. P→C_2 sequence for T-2	04									02				
4. P→C_2 sequence	28	14	05	41	23	10	46	37	19	12	50	40	25	07
1. Term number	43	44	45	46	47	48	49	50	51					
2. P→C_1 sequence for T-1														
3. P→C_2 sequence for T-2														
4. P→C_2 sequence	18	08	45	27	02	38	29	11	47					

FIGURE 37r (2).

(5) Here there are found no contradictions of the nature of those pointed out above. The T-2 matrix appears to have columns of 4 and 5 letters, since the interval between 04 and 02 in row 3 can accommodate a short column of 4 and a long column of 5 letters; the interval between 05 and 04 can accommodate 2 short columns of 4 letters and 1 long column of 5; the intervals between 03 and 06, 06 and 01, 01, and 05 can accommodate long columns of 5 letters each. Only 2 matrices can be constructed of 51 letters with long columns of 5 and short columns of 4 letters. They are:

$$\text{Key of 11}\ \ldots\ldots\ldots \begin{cases} 7\ (\text{long})\ \times 5 = 35 \\ 4\ (\text{short}) \times 4 = \underline{16 \quad 51} \end{cases}$$

$$\text{Key of 12}\ \ldots\ldots\ldots \begin{cases} 3\ (\text{long})\ \times 5 = 15 \\ 9\ (\text{short}) \times 4 = \underline{36 \quad 51} \end{cases}$$

Each of these T-2 matrices is tested as a possibility.

1	2	3	4	5	6	7	8	9	10	11
01	02	03	04	05	06	07	08	09	10	11
12	13	14	15	16	17	18	19	10	21	22
23	24	25	26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41	42	43	44
45	46	47	48	49	50	51				

A

1	2	3	4	5	6	7	8	19	10	11	12
01	02	03	04	05	06	07	08	09	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48
49	50	51									

B

FIGURE 37r (3).

(6) If matrix A is correct, then the numbers in columns 3, 6, 1, 5, 4, and 2 can be transferred to row 3 in figure 37*r* (2); these will permit of inserting numbers in row 2. No contradictions and many checks are found. Here is the diagram:

1. Term number	01	02	03	04	05	06	07	08	09	10	11	12	13	14
2. P→C$_1$ sequence for T–1	04	12	20	28	36	44						42	50	
3. P→C$_1$ sequence for T–2	03	14	25	36	47	06	17	28	39	50	01	12	23	34
4. P→C$_1$ sequence	20	06	48	33	15	44	30	21	03	39	04	42	32	17

1. Term number	15	16	17	18	19	20	21	22	23	24	25	26	27	28
2. P→C$_1$ sequence for T–1	14	22	30						32	40		05	13	21
3. P→C$_1$ sequence for T–2	45	05	16	27	38	49								
4. P→C$_1$ sequence	51	36	22	13	49	31	34	24	09	43	26	16	01	35

1. Term number	29	30	31	32	33	34	35	36	37	38	39	40	41	42
2. P→C$_1$ sequence for T–1						17	25		41	49	03			
3. P→C$_1$ sequence for T–2	04	15	26	37	48					02	13	24	35	46
4. P→C$_1$ sequence	28	14	05	41	23	10	46	37	19	12	50	40	25	07

1. Term number	43	44	45	46	47	48	49	50	51
2. P→C$_1$ sequence for T–1			51	07		23	31	39	
3. P→C$_1$ sequence for T–2									
4. P→C$_1$ sequence	18	08	45	27	02	38	29	11	47

FIGURE 37*r* (4).

(7) In the first place note, in row 2, the constant difference 8, giving many corroborations that the width of the T–1 matrix is 8; in the second place no conflicts whatever become manifest between the pairs of rows. Thus, the validity of the assumption of a T–2 matrix with 11 columns is well established. The rest follows quite readily, with the final result that figure 37*r* becomes completed, and the recovery of both keys is a simple matter. In fact, both keys may be established from a simple study of rows 2 and 3 of the final figure (which would, of course, be identical with that shown in fig. 37*p* and need not here be repeated).

q. A careful study and good grasp of the principles and methods elucidated in this paragraph will be sufficient to indicate to the student that when, as a result of a close study of several messages in the same keys, *partial* C→P sequences become available, the entire C→P sequence or sequences can usually be reconstructed from the partial sequence or sequences and the messages solved without too much difficulty. For instance, suppose it has developed that the enemy has become addicted to stereotypic beginnings, so that the first few letters of a message or of several messages can be reconstructed with some assurance of certainty. The construction of partial C→P sequences and their completion by means of the principles set forth, especially those presented in subparagraphs *m–p*, may result in reconstruction of the complete C→P sequences and ultimate recovery of the transposition key or keys.

28. Special cases of solution of double transposition ciphers.—*a*. When the double transposition system is employed in the field and is used for a voluminous traffic it is almost inevitable that certain situations will arise which make possible a rather easy solution. Aside from the case in which several cryptograms of identical length and in the same key are intercepted, other cases of a special nature may arise. Some of these will be discussed in this paragraph.

b. First, there is the case in which an inexperienced cryptographic clerk fails to execute the double transposition properly and causes the transmission of a cryptogram which is only a single transposition. The solution of this message will be a simple matter and will, of course, yield the key. If the key is the same for both transpositions it is obvious that this will permit the reading of all other messages even though the latter have been correctly cryptographed. The only difficult part of the matter is to find among a large number of intercepted cryptograms one which involves a blunder of this sort. When the cryptanalyst has, as a result of considerable experience, become adept in the solution of transposition ciphers the work of testing cryptograms to ascertain whether or not they involve single columnar transposition is not difficult and goes quite rapidly. For only a few minutes are sufficient to give him the "feeling" that the cryptogram is or is not solvable by single transposition. He might not be able to point out any specific indications which give him this feeling if asked to do so; nevertheless it must be recognized that his intuition is alone sufficient to tell him when there is hope of solution along this line and when further work upon the hypothesis of single transposition is useless.

c. (1) Next comes the case in which the enciphering rectangles of a double transposition cryptogram happen to be perfect squares (that is, both T–1 and T–2 rectangles are perfect squares). In this case, not only is such a cryptogram detectable at once, since the total number of letters is the square of the number of elements in the key, but also the cryptogram can be solved in a very simple manner. For the cryptogram now represents a case in which a completely-filled rectangle has been employed, and moreover there is no need even to assume various widths.

(2) Given the following cryptogram of 49 letters (7×7) as an example, the text is transcribed as shown in figure 39*a* and retranscribed as in figure 39*b*.

Cryptogram.................UCTRN OESHI ETOLR GASOE DUWDD
NOEOE RDNDI RFENC OEEEM NNVE

1	2	3	4	5	6	7
U	S	R	U	O	R	E
C	H	G	W	E	F	E
T	I	A	D	R	E	M
R	E	S	D	D	N	N
N	T	O	N	N	C	N
O	O	E	O	D	O	V
E	L	D	E	I	E	E

FIGURE 39*a*.

1	2	3	4	5	6	7
U	C	T	R	N	O	E
S	H	I	E	T	O	L
R	G	A	S	O	E	D
U	W	D	D	N	O	E
O	E	R	D	N	D	I
R	F	E	N	C	O	E
E	E	M	N	N	V	E

FIGURE 39*b*.

2	6	1	5	3	7	4
C	O	U	N	T	E	R
H	O	S	T	I	L	E
G	E	R	O	A	D	S
W	O	U	N	D	E	D
E	D	O	N	R	I	D
F	O	R	C	E	E	N
E	V	E	N	M	E	N

FIGURE 39*c*.

2	6	1	5	3	7	4
H	O	S	T	I	L	E
F	O	R	C	E	E	N
C	O	U	N	T	E	R
E	D	O	N	R	I	D
G	E	R	O	A	D	S
E	V	E	N	M	E	N
W	O	U	N	D	E	D

FIGURE 39*d*.

(3) The columns of figure 39*b* are now anagrammed, as in figure 39*c*, and the rows rearranged, as in figure 39*d*.

d. When the enciphering rectangle is not a perfect square but nevertheless a complete rectangle, solution of a single cryptogram becomes somewhat more difficult. Here the columns are all equal in length, since the last row of the rectangle is completely filled. Two cases will be considered; first, when the width of the rectangle is a multiple of the depth, or number of letters in the columns, and second, when the depth is a multiple of the width.

e. (1) Taking up the first case, note the following encipherment:

6	2	10	1	7	4	9	8	3	5
W 1	H 2	E 3	N 4	W 5	I 6	L 7	L 8	F 9	I 10
R 11	S 12	T 13	S 14	Q 15	U 16	A 17	D 18	R 19	O 20
N 21	R 22	E 23	A 24	C 25	H 26	G 27	O 28	L 29	D 30
E 31	N 32	V 33	I 34	L 35	L 36	E 37	T 38	O 39	N 40
I 41	G 42	H 43	T 44	A 45	D 46	V 47	I 48	S 49	E 50

T–1 Rectangle

6	2	10	1	7	4	9	8	3	5
N 4	S 14	A 24	I 34	T 44	H 2	S 12	R 22	N 32	G 42
F 9	R 19	L 29	O 39	S 49	I 6	U 16	H 26	L 36	D 46
I 10	O 20	D 30	N 40	E 50	W 1	R 11	N 21	E 31	I 41
W 5	Q 15	C 25	L 35	A 45	L 8	D 18	O 28	T 38	I 48
L 7	A 17	G 27	E 37	V 47	E 3	T 13	E 23	V 33	H 43

T–2 Rectangle

P→C sequence......	34	39	40	35	37	14	19	20	15	17	32	36	31	38	33	2	6	1	8	3	42	46	41	48	43
Cryptogram.........	I	O	N	L	E	S	R	O	Q	A	N	L	E	T	V	H	I	W	L	E	G	D	I	I	H
	4	9	10	5	7	44	49	50	45	47	22	26	21	28	23	12	16	11	18	13	24	29	30	25	27
	N	F	I	W	L	T	S	E	A	V	R	H	N	O	E	S	U	R	D	T	A	L	D	C	G

If the P→C sequence is examined it will be found that sections thereof fall into two categories, as follows:

	Section							Section					
Category A.....	1 —	4	9	10	5	7	Category B....	6 —	2	6	1	8	3
	2 —	14	19	20	15	17		7 —	12	16	11	18	13
	3 —	24	29	30	25	27		8 —	22	26	21	28	23
	4 —	34	39	40	35	37		9 —	32	36	31	38	33
	5 —	44	49	50	45	47		10 —	42	46	41	48	43

(2) There is obviously a definite regularity in the composition of the sections whereby, if the letters corresponding to the numbers in one section can be assembled properly, all the letters corresponding to the numbers in the other sections belonging to the same category (A or B, respectively) will be assembled correctly too. For example, in category B the letters corresponding to the numbers occupying the third, first, and fifth positions in each section are sequent in the plain-text rectangle; in category A the letters corresponding to the numbers occupying the first and fourth positions in each section are sequent. Moreover, all the letters in each section come from the same row in the T-1 rectangle. Consequently, if two sections coming from the same row can be identified, there will be 10 letters which may be rearranged by experiment to form plain text, and the key for this rearrangement will apply to all other pairs of sections. For example, the message in this case has a Q and only one U. The Q (P→C sequence No. 15) is in the second section, the U (P→C sequence No. 16) is in the seventh section. These two sections come from the same row and the letters may be anagrammed:[4]

1	2	3	4	5		6	7	8	9	10
S	R	O	Q	A	and	S	U	R	D	T

2 or S	1 or 6	10	6 or 1	4	7	5	9	8 or 2	3
R	S	T	S	Q	U	A	D	R	O

Experiment may now be made with two other sections, applying the same transposition. Thus:

1	2	3	4	5		·6	7	8	9	10
I	O	N	L	E	and	N	L	E	T	V

2 or 8	1 or 6	10	6 or 1	4	7	5	9	8 or 2	3
O	I		N					E	
E	N	V	I	L	L	E	T	O	N

Obviously the proper key for rearrangement is 8-6-10-1-4-7-5-9-2-3. By continuing this procedure the following additional rows of the T-1 rectangle are reconstructed.

1	2	3	4	5		6	7	8	9	10		8	6	10	1	4	7	5	9	2	3
N	F	I	W	L	and	H	I	W	L	E	yields......	W	H	E	N	W	I	L	L	F	I
T	S	E	A	V	and	G	D	I	I	H	yields......	I	G	H	T	A	D	V	I	S	E
A	L	D	C	G	and	R	H	N	O	E	yields......	N	R	E	A	C	H	G	O	L	D

---

[4] The fact that the length of the sections corresponds to 5-letter groups has, of course, no bearing on the validity of the method. In this case it just happens that the rectangle contains 5 letters per column.

The various rows are now assembled in sequence, giving the following:

```
W H E N W I L L F I
R S T S Q U A D R O
N R E A C H G O L D
E N V I L L E T O N
I G H T A D V I S E
```

The transposition key can now be reconstructed with ease.

(3) The cryptanalyst in this case must, of course, make an assumption as to the width of the enciphering rectangle before he can apply the method. With a number such as 50, the dimensions 10×5 or 5×10 suggest themselves. The process of finding cipher groups which form pairs on the same row is one of "cut and try." If there is a single Q and a single U in the message, the initial pair of groups is obvious.

f. When the depth of the rectangle is a multiple of the width, solution follows along the lines of the preceding case. Taking the same message as before, note what happens in encipherment with a rectangle of 5 columns containing 10 letters each:

2	5	1	4	3
1 W	2 H	3 E	4 N	5 W
6 I	7 L	8 L	9 F	10 I
11 R	12 S	13 T	14 S	15 Q
16 U	17 A	18 D	19 R	20 O
21 N	22 R	23 E	24 A	25 C
26 H	27 G	28 O	29 L	30 D
31 E	32 N	33 V	34 I	35 L
36 L	37 E	38 T	39 O	40 N
41 I	42 G	43 H	44 T	45 A
46 D	47 V	48 I	49 S	50 E

T-1

2	5	1	4	3
3 E	8 L	13 T	18 D	23 E
28 O	33 V	38 T	43 H	48 I
1 W	6 I	11 R	16 U	21 N
26 H	31 E	36 L	41 I	46 D
5 W	10 I	15 Q	20 O	25 C
30 D	35 L	40 N	45 A	50 E
4 N	9 F	14 S	19 R	24 A
29 L	34 I	39 O	44 T	49 S
2 H	7 L	12 S	17 A	22 R
27 G	32 N	37 E	42 G	47 V

T-2

P→C sequence---13 38 11 36 15 40 14 39 12 37 3 28 1 26 5 30 4 29 2 27 23 48 21 46 25 50 24 49 22 47
Cryptogram-----T T R L Q N S O S E E O W H W D N L H G E I N D C E A S R V

18 43 16 41 20 45 19 44 17 42 8 33 6 31 10 35 9 34 7 32
D H U I O A R T A G L V I E I L F I L N

Taking the numbers of the P→C sequence and arranging them in sections of 10, the results are as follows:

1	2	3	4	5	6	7	8	9	10
3	28	1	26	5	30	4	29	2	27
8	33	6	31	10	35	9	34	7	32
13	38	11	36	15	40	14	39	12	37
18	43	16	41	20	45	19	44	17	42
23	48	21	46	25	50	24	49	22	47

It is obvious that if the 3d, 9th, 1st, 7th, and 5th columns are made sequent, good text will be produced within the 5 rows. Thus:

1	2	3	4	5	6	7	8	9	10
T	T	R	L	Q	N	S	O	S	E
E	O	W	H	W	D	N	L	H	G
E	I	N	D	C	E	A	S	R	V
D	H	U	I	O	A	R	T	A	G
L	V	I	E	I	L	F	I	L	N

3	9	1	7	5
R	S	T	S	Q
W	H	E	N	W
N	R	E	A	C
U	A	D	R	O
I	L	L	F	I

The subsequent steps are obvious. Here again in solving an unknown example it would be necessary to test out various assumptions with respect to the dimensions of the rectangle before attempting to apply the method outlined.

g. Whenever this simple relationship between the width and depth of the rectangle obtains, that is, when one dimension is a multiple of the other, solution of a single cryptogram is relatively easy. The reason for this is not hard to see. When the enciphering rectangle is a perfect square, every column of the T–2 rectangle is composed of letters which all come from the same row of the T–1 rectangle. Hence solution is in this case the same as though a false double transposition were in effect, with merely the columns and the rows of a single rectangle shifted about. When the width of the transposition rectangle is twice the depth, a column of the T–2 rectangle contains half the letters appearing on one row of the T–1 rectangle; two columns therefore contain all the letters belonging in the same row of the T–1 rectangle. If the width were three times the depth, then three columns of the T–2 rectangle would contain all the letters belonging in the same row of the T–1 rectangle, and so on. When the width is half the depth, a column of the T–2 rectangle contains all the letters appearing in two rows of the T–1 rectangle; when the width is one-third the depth, a column of the T–2 rectangle contains all the letters appearing in three rows of the T–1 rectangle, and so on. But when this multiple relationship no longer obtains, solution becomes more difficult because each column of the T–2 rectangle is composed of letters coming from several columns of the T–1 rectangle, in an irregular distribution. Solution is, of course, most difficult when incompletely filled rectangles are used. However, although solvable, even in the case of a single message, the solution will not be dealt with in this text.

SECTION VI

PRINCIPLES OF MATRIX RECONSTRUCTION

Paragraph
Special designs or geometric figures.. 29
Reconstruction of transposition matrix.. 30

29. Special designs or geometric figures.—*a.* It is impossible here to elucidate and demonstrate by example all the methods which may be used for the solution of cryptograms produced by the many various types of transposition designs or geometric figures other than the simple rectangular ones thus far treated. Reference may be made to such matrices as triangles, trapezoids, and polygons of various symmetrical shapes. Most of these matrices, however, are impractical for military correspondence in any case, so that no attention need be given them in this text.

b. If such designs were used, although it might be difficult to solve a single or even a few messages in the same key, the general solution described in paragraph 26 is applicable whenever two or more messages of identical lengths but in the same key are available for study. Since most of these designs are of a fixed or inflexible character with regard to the number of letters that can be accommodated with one application of the design to the plain text to be enciphered, the production of several cryptograms of identical length in the same key is by no means an unusual circumstance. The general solution can usually be depended upon to yield the answer to cryptograms of this category but it then becomes advisable to try to ascertain the exact nature of the specific design or geometric figure employed, that is, to reconstruct the transposition matrix. For this purpose a general method will be indicated by means of a specific example, leaving other cases to the ingenuity of the student after he has learned the general method.

30. Reconstruction of transposition matrix.—*a.* Assume that the enemy is employing an unknown geometric figure of rather small dimensions so that it appears from a study of the traffic that it accommodates a maximum of 85 letters. A long cryptogram has been intercepted and it is broken up into sections of 85 letters, which sections are then superimposed, as shown below. It will be noted that there are 3 complete sections of 85 letters each, plus a final section of but 49 letters. The final section will be dealt with later.

```
   1 2 3 4 5   6 7 8 9 10   11 12 13 14 15   16 17 18 19 20   21 22 23 24 25   26 27 28 29 30
1  T D N F R   A O I S J    F  E  R  O  E    E  A  R  Y  O    I  E  P  T  L    T  H  A  V  N
2  W R T D L   U O S F C    N  N  T  U  I    N  M  O  S  X    L  N  O  N  P    A  T  S  I  F
3  M A I S V   I T S O T    H  L  T  E  S    R  I  O  V  I    Y  V  W  N  G    P  E  O  O  I
4  G R U T S   O E B R M    L  R  M  O  O    E  T  C  N  N    D  Y  E  E  H    T  Q  C  N  T

   31 32 33 34 35   36 37 38 39 40   41 42 43 44 45   46 47 48 49 50   51 52 53 54 55   56 57 58 59 60
1  A  N  N  C  T    S  Y  O  A  A    C  E  M  E  H    I  E  I  B  I    H  A  D  E  X    T  C  T  U  R
2  W  D  H  E  B    R  N  D  T  T    D  I  Y  A  F    A  D  A  G  R    D  O  E  O  A    A  J  T  R  E
3  A  T  U  A  C    O  D  P  O  B    I  M  N  R  T    I  N  E  S  H    O  Y  N  F  L    I  H  N  R  O
4  M  O  O  C  E    O  I  B  R  S    E  P  Y  C  S    S  S  S  F

   61 62 63 64 65   66 67 68 69 70   71 72 73 74 75   76 77 78 79 80   81 82 83 84 85
1  R  F  V  S  T    N  N  R  U  I    N  O  U  R  T    F  F  E  N  V    E  L  N  O  E
2  M  A  I  O  N    V  O  T  O  T    T  R  N  O  I    E  U  A  N  H    R  O  C  T  A
3  L  T  Y  E  X    A  E  U  O  A    E  F  R  T  E    X  Y  R  V  R    A  U  I  N  T
```

b. The anagramming process is applied to the superimposed complete sections, using the letter J in the first section as a starting point and building up text on either side, until the following partially reconstructed text is obtained:

40	34	45	85	2	61	20	28	53	10	69	79	41	35	46	84	3	62	21	29	54	11
A	C	H	E	D	R	O	A	D	J	U	N	C	T	I	O	N	F	I	V	E	F
T	E	F	A	R	M	X	S	E	C	O	N	D	B	A	T	T	A	L	I	O	N
B	A	T	T	A	L	I	O	N	T	O	V	I	C	I	N	I	T	Y	O	F	H

c. Examining the numbers forming this partial C→P sequence, note the following sections of the sequence:

40 34 45 85 2 61 20 28 53 10 . . .
41 35 46 84 3 62 21 29 54 11

They show a quite definite relationship, leading to the suspicion that the C→P sequence is systematic in its composition. The numbers are then written down on cross-section paper so that consecutive numbers appear on the same level, as shown in figure 40–A:

C→P sequence		40	34	45	85	2	61	20	28	53	10	69	79	41	35	46	84	3	62	21	29	54	11
		A	C	H	E	D	R	O	A	D	J	U	N	C	T	I	O	N	F	I	V	E	F
Text		T	E	F	A	R	M	X	S	E	C	O	N	D	B	A	T	T	A	L	I	O	N
		B	A	T	T	A	L	I	O	N	T	O	V	I	C	I	N	I	T	Y	O	F	H
	1										10												11
	2									53		69										54	
	3								28				79								29		
Matrix reconstruction diagram	4	40						20						41						21			
	5		34				61								35				62				
	6			45		2										46		3					
	7				85												84						

FIGURE 40–A.

d. From this skeleton of what may be termed the *matrix-reconstruction diagram* it is possible to derive direct clues for the continuance and completion of the C→P sequence and the text of the message. For example, it would appear that the very next column to the left should be 78, the one to the left of 78 should be 68, the one to the left of 68 should be 9. Trial gives the following:

9	68	78	40	34	45	85	2	. . .
S	R	E	A	C	H	E	D	. . .
F	T	A	T	E	F	A	R	. . .
O	U	R	B	A	T	T	A	. . .

To the right of column 11 should come columns 70 80 42 36 47. Thus:

29	54	11	70	80	42	36	47	. . .
V	E	F	I	V	E	S	E	. . .
I	O	N	T	H	I	R	D	. . .
O	F	H	A	R	M	O	N	. . .

This, of course, speeds up the work involved in the anagramming process and when completed the text, the C→P sequence, the P→C sequence, and the matrix reconstruction diagram are as shown in figure 40–B. In the cells of the diagram there have been inserted in the upper left hand corner small numbers in italics, the latter numbers being merely the term numbers applying to the C→P sequence.

e. The matrix-reconstruction diagram in figure 40–B shows a total of 7 levels of numbers. Let the term numbers corresponding to the *consecutive* C→P sequence numbers on the same level in the diagram be set down. Thus, for the C→P sequence numbers 4 to 16, inclusive, on the first level the term numbers are:

C→P sequence number	4	5	6	7	8	9	10	11	12	13	14	15	16
Term number	1	3	7	13	21	31	43	55	65	73	79	83	85

On the second level there are two sets of consecutive C→P sequence numbers, those from 48 to 58, inclusive forming one set, those from 64 to 74, inclusive forming the other set. Two series of term numbers are therefore derived:

C→P sequence number	48	49	50	51	52	53	54	55	56	57	58
Term number	2	6	12	20	30	42	54	64	72	78	82

C→P sequence number	64	65	66	67	68	69	70	71	72	73	74
Term number	4	8	14	22	32	44	56	66	74	80	84

What has been said of the 2d level applies also to the remaining levels, and the term numbers are therefore set down in the following tabular form:

	1	2	3	4	5	6	7	8	9	10	11	12	13	C→P sequence numbers to which applicable
1	1	3	7	13	21	31	43	55	65	73	79	83	85	(4–16)
2	2	6	12	20	30	42	54	64	72	78	82			(48–58)
3	4	8	14	22	32	44	56	66	74	80	84			(64–74)
4	5	11	19	29	41	53	63	71	77					(24–32)
5	9	15	23	33	45	57	67	75	81					(75–83)
6	10	18	28	40	52	62	70							(17–23)
7	16	24	34	46	58	68	76							(38–44)
8	17	27	39	51	61									(59–63)
9	25	35	47	59	69									(33–37)
10	26	38	50											(1–3)
11	36	48	60											(45–47)
12	37													(85)
13	49													(84)

FIGURE 41

f. There are in all 13 sets or series of consecutive C→P sequence numbers, indicating that the transposition matrix has 13 columns, the number of letters in each column corresponding with the number of different terms in each series. Thus, there is a column of 13 letters, 2 columns of 11 letters, 2 columns of 9 letters, and so on. This leads directly to the idea of a very symmetrical matrix of the form shown in figure 42–A.

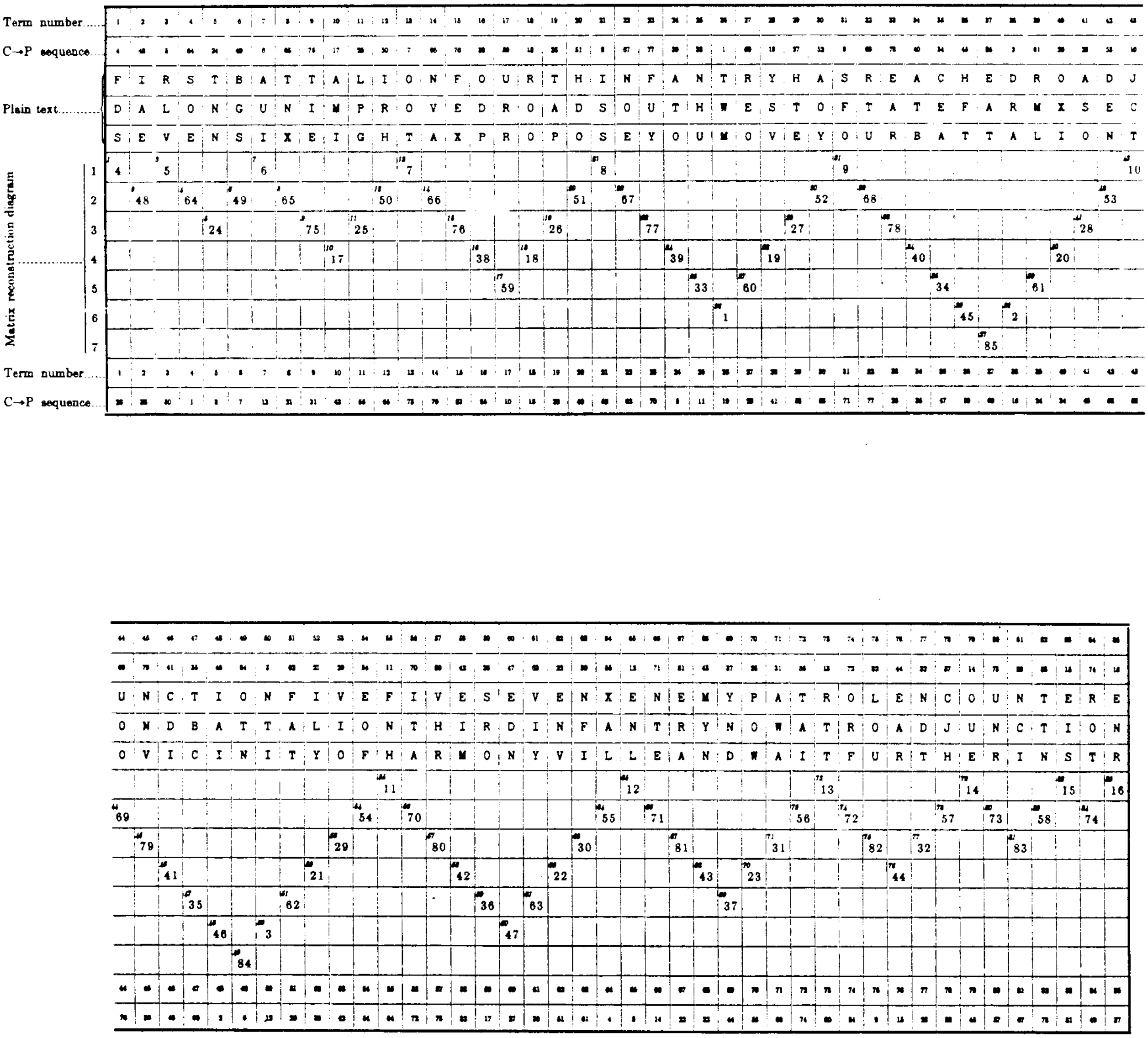
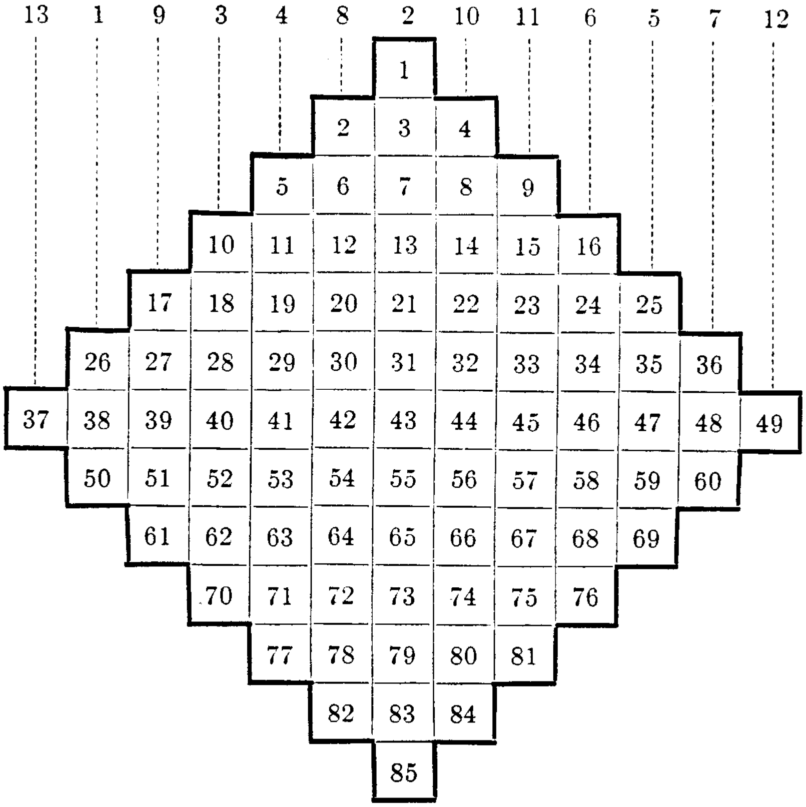

Term number
C→P sequence
Plain text
F I R S T B A T T A L I O N F O U R T H I N F A N T R Y H A S R E A C H E D R O A D J
D A L O N G U N I M P R O V E D R O A D S O U T H W E S T O F T A T E F A R M X S E C
S E V E N S I X E I G H T A X P R O P O S E Y O U M O V E Y O U R B A T T A L I O N T
Matrix reconstruction diagram
Term number
C→P sequence
U N C T I O N F I V E F I V E S E V E N X E N E M Y P A T R O L E N C O U N T E R E
O N D B A T T A L I O N T H I R D I N F A N T R Y N O W A T R O A D J U N C T I O N
O V I C I N I T Y O F H A R M O N Y V I L L E A N D W A I T F U R T H E R I N S T R

Figure 40-B

						1						
					2	3	4					
				5	6	7	8	9				
			10	11	12	13	14	15	16			
		17	18	19	20	21	22	23	24	25		
	26	27	28	29	30	31	32	33	34	35	36	
37	38	39	40	41	42	43	44	45	46	47	48	49
	50	51	52	53	54	55	56	57	58	59	60	
		61	62	63	64	65	66	67	68	69		
			70	71	72	73	74	75	76			
				77	78	79	80	81				
					82	83	84					
						85						

FIGURE 42-A.

g. The recovery of the transposition key (for the columns of figure 42-A) is now a simple matter. Referring to the P→C sequence shown in figure 40-B, and noting the various columns in figure 42-A in which successive numbers of the P→C sequence fall, the key number 1 of the transposition key obviously applies to the column containing P→C sequence numbers 26-38-50; the key number 2 obviously applies to the column containing P→C sequence numbers 1-3-7-13-21-31-43-55-65-73-79-83-85; and so on. The complete transposition key and the matrix are shown in figure 42-B.



13	1	9	3	4	8	2	10	11	6	5	7	12
						1						
					2	3	4					
				5	6	7	8	9				
			10	11	12	13	14	15	16			
		17	18	19	20	21	22	23	24	25		
	26	27	28	29	30	31	32	33	34	35	36	
37	38	39	40	41	42	43	44	45	46	47	48	49
	50	51	52	53	54	55	56	57	58	59	60	
		61	62	63	64	65	66	67	68	69		
			70	71	72	73	74	75	76			
				77	78	79	80	81				
					82	83	84					
						85						

FIGURE 42-B.

h. The solution of the incomplete section of the message (the last 79 letters) now becomes a simple matter, since the matrix and the transposition key are both known. The matter can be handled as if simple transposition were involved, by outlining the matrix to contain exactly 79 letters and inscribing the cipher text in the columns in key-number order.

i. The foregoing principles and procedure will be found quite valuable not only in facilitating the anagramming of the text in its initial stages (as in subparagraph *d*) but also in reconstructing various types of matrices based upon symmetrical designs used with single transposition (subparagraphs *e–g*). It should be noted that the number of *levels* in the reconstruction diagram corresponds with the number of different-length columns in the matrix; the number of different *categories* of term numbers (as in figure 41) corresponds with the number of columns in the matrix.

SECTION VII

SOLUTION OF GRILLES

Paragraph
Revolving grilles ---------- 31
Solution of example ---------- 32
Concluding remarks on the solution of revolving grilles ---------- 33
Indefinite or continuous grilles ---------- 34

31. Revolving grilles.—*a.* In this type of grille [1] apertures are distributed among the cells of a square sheet of cross-section paper in such a manner that when the grille is placed upon a grid (a sheet of cross-section paper of the same size as the grille) certain cells of the grid are disclosed; then when the grille is turned three times successively through angles of 90° from an initial position upon the grid, all the remaining undisclosed grid cells (or all but the central grid cell) are disclosed in turn. Correspondents must, of course, possess identical grilles and they must have an understanding as to its initial position and direction of rotation, clockwise or counterclockwise. There are two procedures possible in using such a grille. (1) The letters of the plain text may be inscribed successively in the grid cells through the apertures of the grille; when the grid has been completely filled the grille is removed and the letters transcribed from the grid according to a prearranged route. (2) All the letters of the plain text may first be inscribed in the grid cells according to a prearranged route and then the grille applied to the completely-filled grid to give the sequence of letters forming the cryptogram. The two methods of using the grille are reciprocal; if the first-described method is used to encipher a message, the second is used to decipher the cryptogram, and vice versa. The first of the two above-described methods, the one in which the plain text is inscribed through the apertures, will here be referred to as the *alpha* method; the second method will be referred to as the *beta* method.

b. The number of letters in a cryptogram enciphered by such a device is either a perfect square, when the grille has an even number of cells per side, or is 1 less than a perfect square, when the grille has an odd number of cells per side, in which case the central cell of the grid is not disclosed and hence remains unfilled.[2]

c. The manner of construction and the method of use of a grille entails certain consequences which can be employed to solve the cryptograms and to reconstruct the grille itself. The student who wishes to get a thorough grasp of the underlying principles to be explained will do well to prepare a grille [3] and study the properties which characterize cryptograms produced by its use. Three principles will be brought to bear in the solution of grille ciphers of this type and they will be demonstrated by reference to the grille and message shown in figure 43.

---

[1] See Special Text No. 166, *Advanced Military Cryptography*, sec. V.

[2] Of course, the cryptogram may consist of the letters produced by several applications of the same grille. For example, if a message of 170 letters is to be enciphered by a grille accommodating only 36 letters at a time, the message is divided up into 5 sections of 36 letters each (10 nulls being added to make the total a multiple of 36). The total number of letters (180) here shows no properties of the type noted. Again, if the grille has a capacity greater than the number of letters to be enciphered, certain of the grid cells may be cancelled, so that the number of letters in the final cryptogram will not be a perfect square or 1 less than a perfect square.

[3] Detailed instructions for the construction of revolving grilles will be found in Special Text No. 166, *Advanced Military Cryptography*, sec. V.

MESSAGE

YOUR LINES TO THIS COMMAND POST CUT BY SHELL FIRE REQUEST YOU CHANGE THE ROUTE.

Grille: 8×8.

a b

	1	2	3	4	5	6	7	8
1	1	2	3	☒	5	6	7	☒
2	7	☒	2	3	4	☒	1	2
3	6	5	1	2	3	☒	2	3
4	5	☒	3	☒	1	2	3	4
5	4	3	2	1	1	3	4	☒
6	☒	2	1	☒	☒	1	5	☒
7	☒	1	5	4	☒	☒	1	7
8	1	☒	6	5	4	3	2	1

d c

A

POSITION 1

a b

1	2	3	Y 4	5	6	7	O 8
9	U 10	11	12	13	R 14	15	16
17	18	19	20	21	L 22	23	24
25	I 26	27	N 28	29	30	31	32
33	34	35	36	37	38	39	E 40
S 41	42	43	T 44	O 45	46	47	T 48
H 49	50	51	52	I 53	S 54	55	56
57	C 58	59	60	61	62	63	64

d c

B

POSITION 2

d a

1	O 2	M 3	4	5	6	7	8
M 9	10	11	12	A 13	14	N 15	16
17	18	19	20	21	22	23	24
25	26	D 27	28	P 29	30	31	O 32
33	S 34	T 35	36	37	38	39	40
41	C 42	43	44	45	U 46	T 47	48
49	50	51	52	53	54	55	56
57	58	B 59	Y 60	61	62	63	S 64

c b

C

FIGURE 43

D

POSITION 3

ɔ p

1	2	3	4	5	6	H 7	8
9	10	E 11	L 12	13	14	15	L 16
F 17	18	19	I 20	R 21	22	23	E 24
R 25	26	27	28	29	30	31	32
33	34	35	36	E 37	38	Q 39	40
41	42	U 43	44	45	46	47	48
49	50	E 51	52	53	54	S 55	56
T 57	58	59	60	Y 61	62	63	64

q ɐ

E

POSITION 4

b c

O 1	2	3	4	U 5	C 6	7	8
9	10	11	12	13	14	15	16
17	H 18	A 19	20	21	22	N 23	24
25	26	27	28	29	G 30	E 31	32
T 33	34	35	H 36	37	E 38	39	40
41	42	43	44	45	46	47	48
49	R 50	51	O 52	53	54	55	U 56
57	58	59	60	61	T 62	E 63	64

a d

FINAL GRID

a b

O 1	O 2	M 3	Y 4	U 5	C 6	H 7	O 8
M 9	U 10	E 11	L 12	A 13	R 14	N 15	L 16
F 17	H 18	A 19	I 20	R 21	L 22	N 23	E 24
R 25	I 26	D 27	N 28	P 29	G 30	E 31	O 32
T 33	S 34	T 35	H 36	E 37	E 38	Q 39	E 40
S 41	C 42	U 43	T 44	O 45	U 46	T 47	T 48
H 49	R 50	E 51	O 52	I 53	S 54	S 55	U 56
T 57	C 58	B 59	Y 60	Y 61	T 62	E 63	S 64

F

d c

CRYPTOGRAM

```
OOMYU CHOMU ELARN LFHAI RLNER IDNPG
EOTST HEEQE SCUTO UTTHR EOISS UTCBY
YTES
```

FIGURE 43—Continued.

d. The first principle may be termed that of *symmetry*. When a revolving grille is in position 1 a certain number of cells of the underlying grid are disclosed (uncovered). For each such disclosed cell of the grid there is a symmetrically-corresponding cell on the same grid which is disclosed when the grille is turned to positions 2, 3, and 4, because the apertures of the grille remain fixed—only their positions change as the grille is turned in the process of encipherment. Now two *successive* apertures in position 1 will, of course, be occupied by a plain-text digraph (*alpha* method of encipherment). When the grille reaches position 3, after a turn of 180°, the two apertures concerned will disclose two cells which will also be occupied by a plain-text digraph, *but the letters composing the digraph will be in reverse order in the plain text.* This property is true also of two *successive* apertures in position 2 when they turn up in position 4. Let the student verify this by means of the grille which he has constructed. Thus, referring to figure 43, at *A* is shown the grille in position 1. In the first row are shown 2 apertures, at coordinates 1–4 and 1–8. At *B* are shown the results of the first application of the grille to the grid. Note the letters Y O (first 2 letters of message) in cells 4 and 8. Now note that the symmetrically-corresponding cells disclosed when the grille is in position 3 are cells 57 and 61 and these correspond to cells 4 and 8 in the reverse order. The letter T in cell 57 therefore symmetrically corresponds with letter O in cell 8; the letter Y in cell 61 corresponds with letter Y in cell 4. The same is true of all other letters in positions 1 and 3. As a consequence of this property of grilles, a single cryptogram can be handled as though it were really two cryptograms of identical length having certain characteristics by means of which an assumption made in one text may be verified by what it yields in the other text. That is, when the cryptogram is transcribed as a series of letters in one line and the same text is written in another line under these letters but in reversed order, then the superimposed letters will bear the symmetrical relationship pointed out in this paragraph. If two letters in the upper line of such a transcription are taken to form a digraph, the two corresponding letters in the lower line must form a digraph but in reversed order in the plain text. For example, if the cryptogram of figure 43 is written out as explained above, the result is as shown at figure 44. Now the presence of the Q in position 39 suggests that it be combined with a U. If the U in position 43 is taken, then the symmetrical digraph corresponding to Q U would be L I; if the U in position 56 is taken, the symmetrically-corresponding digraph would be M I. Furthermore, two apertures which are in the same column *and which do not have an intervening aperture between them,* will yield a good digraph in all 4 positions of the grille. For example, note apertures 2–6 and 3–6 at A in figure 43. When the grille is turned to positions 2, 3, and 4 they will disclose two sequent letters in each case. An analysis of the symmetries produced by an 8×8 grille yields the following table, which shows what cells are disclosed in the other 3 positions when an aperture is cut in any one cell in 1 of the 4 positions of the grille. For example, an aperture cut in cell 11 (position 1) will disclose grid-cell 23 when the grille takes position 2, grid-cell 54 when the grille takes position 3, and grid-cell 42 when the grille takes position 4.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
O	O	M	Y	U	C	H	O	M	U	E	L	A	R	N	L	F	H	A	I	R	L	N	E	R	I	D	N	P	G	E	O
S	E	T	Y	Y	B	C	T	U	S	S	I	O	E	R	H	T	T	U	O	T	U	C	S	E	Q	E	E	H	T	S	T

33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
T	S	T	H	E	E	Q	E	S	C	U	T	O	U	T	T	H	R	E	O	I	S	S	U	T	C	B	Y	Y	T	E	S
O	E	G	P	N	D	I	R	E	N	L	R	I	A	H	F	L	N	R	A	L	E	U	M	O	H	C	U	Y	M	O	O

FIGURE 44.

Positions:	1-3	2-4	1-3	2-4	1-3	2-4	1-3	2-4
	1	8	5	25	11	23	19	22
	64	57	60	40	54	42	46	43
	2	16	6	17	12	31	20	30
	63	49	59	48	53	34	45	35
	3	24	7	9	13	26	21	27
	62	41	58	56	52	39	44	38
	4	32	10	15	14	18	28	29
	61	33	55	50	51	47	37	36

FIGURE 45.

The second principle may be termed that of *exclusion*. On account of the system upon which the construction of a revolving grille is based, a knowledge of the location of an aperture in one of the bands brings with it a knowledge of 3 other locations in which there can be no apertures. For example, referring to A in figure 43, the presence of the aperture at coordinates 1–4 precludes the presence of apertures at coordinates 4–8, 8–5, and 5–1. By virtue of this principle of exclusion, the number of possibilities for choice of letters in solving a cryptogram prepared by means of a revolving grille becomes much reduced and the problem is correspondingly simplified, as will be seen presently.

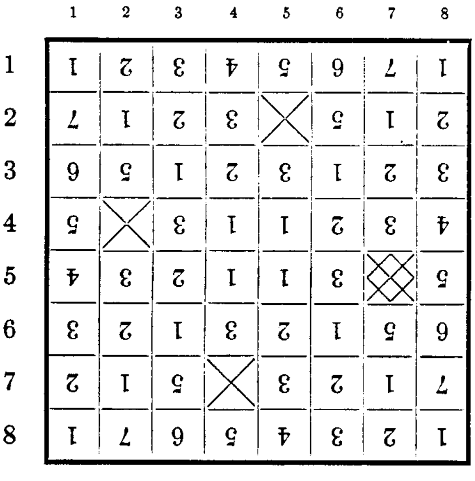
f. The third principle may be termed that of *sequence*. When trying to build up text, the letters which follow a given sequence of plain-text letters will usually be found to the right and below, that is, if the normal method of writing was used (left to right and from the top downward). For example, referring to figure 44, if the trigraph Q U E is to be built up, neither the U in position 5 nor the U in position 10 is very likely to be the one that follows the Q; the U in position 43 is the most likely candidate because it is the first one beyond the Q. Suppose the U in position 43 is selected. Then the E for Q U E cannot be the one in position 40, or in any position in front of 40, since the E must be beyond the U in the diagram.

g. In solving a grille, it will be found advisable to prepare a piece of cross-section paper of proper size for the grille and to cut each aperture as soon as its location in the grille becomes quite definite. In this way not only will the problem be simplified but also when completed the proper grille is at hand.

32. Solution of example.—*a*. Suppose the cryptogram shown in figure 43 is to be solved. It has 64 letters, suggesting a grille 8×8. The cryptogram is first transcribed into a square 8×8, yielding what has already been obtained at F in figure 43. The Q in position 39 suggests that it is part of a word inscribed when the grille was in position 3, since there will be 16 plain-text letters inscribed at each position of the grille. Then a piece of cross-section paper is prepared for making the grille as shown in figure 45–A, and an aperture is cut in the proper position to disclose, in position 3, cell 39. It will be found that this is the aperture located at coordinates 4–2 of the grille shown in figure 45–A. At the same time the other 3 cells numbered 4 in the second band of the grille are marked so that they cannot become apertures. The result is shown in figure 45–B. Conforming to the principle of sequence, the U to be combined with the Q is sought to the right of the Q in figure 43–F. There are three candidates, in positions 43, 46, and 56. They yield:

	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	1
2	7	1	2	3	4	5	1	2
3	6	5	1	2	3	1	2	3
4	5	4	3	1	1	2	3	4
5	4	3	2	1	1	3	4	5
6	3	2	1	3	2	1	5	6
7	2	1	5	4	3	2	1	7
8	1	7	6	5	4	3	2	1

FIGURE 45–A.



	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	1
2	7	1	2	3	X	5	1	2
3	6	5	1	2	3	1	2	3
4	5	X	3	1	1	2	3	4
5	4	3	2	1	1	3	X	5
6	3	2	1	3	2	1	5	6
7	2	1	5	X	3	2	1	7
8	1	7	6	5	4	3	2	1

FIGURE 45–B.

(Grille in position 3)

39	43	39	46	39	56
Q	U	Q	U	Q	U
I	L(=L I_D)	I	A(=A I_D)	I	M(=M I_D)

All of the symmetrical correspondents of these 3 Q U's are good digraphs and it is impossible to eliminate any of the three possibilities. The U in position 43 would place an aperture at coordinates 6–3 in figure 45–B; the U in position 46 would place an aperture at coordinates 6–6; and the U in position 56 would place an aperture at coordinates 7–8. All of these are possible, none being excluded by principle 2. Suppose the Q U is followed by E. There are only two possibilities: an E in cell 51 and E in cell 63. The following possibilities are presented:

39	43	51	39	43	63	39	46	51	39	46	63	39	56	63
Q	U	E	Q	U	E	Q	U	E	Q	U	E	Q	U	E
I	L	R	I	L	O	I	A	R	I	A	O	I	M	O
(=R	L	I)	(=O	L	I)	(=R	A	I)	(=O	A	I)	(=O	M	I)

None of the symmetrical correspondents of the Q U E's are impossible sequences in plain text, although O A I is not as probable as the others. (The O could be the end of a word, the AI the beginning of the word AID, AIM, AIR, etc.) Each of these possibilities would be tested by principle 2 to see if any conflicts would arise as to the positions of apertures. As in all cases of transposition ciphers, the most difficult part of the solution is that of forcing an entering wedge into the structure and getting a good start; when this has been done the rest is easy. Note what the results are when the proper apertures are assumed for QUEST in this case, as shown in figure 45–C. In position 1 this yields OUR LI . . .; in position 2 it yields two digraphs AN and UT; in position 4 it yields two digraphs H A and R O. The student should note that the indicated digraphs A N and R O in positions 2 and 4, respectively, are certain despite the fact that there is a space between the two apertures disclosing these letters, for the principle of exclusion has permitted the crossing off of this cell as a possibility for an aperture.

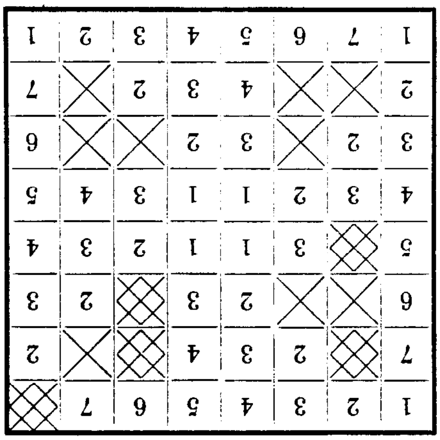


FIGURE 45-C.
(Grille in position 3)

b. Enough has been shown of the procedure to make further demonstration unnecessary. Given the sequence OUR LI one begins to build on that, assuming a word such as LINE. This yields possibilities for the placement of additional apertures in the grille; these are tested in positions 2, 3, 4, and so on. When any 16 consecutive letters of plain text have been established all apertures have been ascertained and the problem has been completed. Subsequent cryptograms prepared by the same grille can be read at once.

c. If attempts at solution on the basis of the alpha method of using a grille have failed, the obvious modifications in procedure on the basis of the beta method can readily be made.

33. Concluding remarks on the solution of revolving grilles.—*a.* There is nothing about the mechanics of revolving grilles which prevents their employment in enciphering complete words instead of individual letters. However, the assembling of whole words in intelligible sequences and thus the reconstruction of the original plain text is a much easier matter than assembling single letters to form the words of the original plain text

b. In case the same grille has been employed several times with separate grids to encipher a message that is considerably longer than a single grid will accommodate (see footnote 2, par. 31*b*), the several sections each representing the set of letters enciphered on one grid may be superimposed and the general solution described in paragraph 26 may then be applied.

c. In case the capacity of a grille is in excess of the number required by the length of the text to be enciphered, either of two procedures may be agreed upon. The grid cells which would otherwise be unoccupied may be filled by nulls, or the grid may be left incomplete. As regards the former procedure, little more need be said than that the presence of a few nulls will only delay solution a bit until the fact that nulls are being employed for this purpose becomes established. But the second type of procedure calls for more comment. If the grid is to be left incomplete it is necessary, before applying the grille, to count the number of plain-text letters and to cancel from the grid a number of cells equal to the number of cells in excess of the total number required. The position of the cells to be cancelled must be agreed upon: commonly, they are those at the end of the grid. Such cells are marked so that when they become exposed during the rotations of the grille they will not be used. Thus, for example, the grille shown in figure 43–A is intended for a grid of 64 letters; if the message to be enciphered contains only 53 letters, 12 cells of the grid must be canceled, and by agreement they may be cells 53 to 64, inclusive. The solution of a single cryptogram of this sort, or even of several of them of different lengths, may become a rather difficult matter. First of all, clues as to the dimensions of the grille are no longer afforded by the total number of letters in the cryptogram, so that this information can be obtained only by more or less laborious experimentation. Grilles of various dimensions must be assumed, one after the other, until the correct dimensions have been found. In the second place, the symmetrical relationships pointed out in paragraph 31 no longer obtain, so that a single cryptogram cannot be handled as though it were constituted of two messages of identical length. Of course, in trying out any assumed dimensions, the 64 letters of the cryptogram may be written out in two superimposed lines, blanks being left for those positions which are unfilled. The procedure then follows the normal lines. About the most hopeful clues would be obtained from a knowledge of the circumstances surrounding the transmission and affording a basis for the assumption of probable words. However, were such a system employed for regular communication there would undoubtedly be cases of cryptograms of identical lengths, so that the type of solution given in paragraph 26 will be applicable. Once a solution of this sort has been obtained, the dimensions of the grille may be ascertained. Subsequent cryptograms may then be attacked on the basis of the normal procedure, with such modifications as are indicated by the absence of the number of letters needed to make a completely-filled grid.

34. Indefinite or continuous grilles.—*a.* In his *Manual of Cryptography,* Sacco illustrates a type of grille which he has devised and which has elements of practical importance. An example of such a grille is shown in figure 46. This grille contains 20 columns of cells, and each column contains 5 apertures distributed at random in the column. There are therefore 100 apertures in all, and this is the maximum number of letters which may be enciphered in one position of the grille. The plain text is inscribed *vertically,* from left to right, using only as many columns as may be necessary to inscribe the complete message. A 25-letter message would require but 5 columns. To form the cryptogram the letters are transcribed *horizontally* from the rows, taking the letters from left to right as they appear in the apertures. If the total number of letters is not a multiple of 5, sufficient nulls are added to make it so. In decryptographing, the total number of letters is divided by 5, this giving the number of columns employed. The cipher text is inscribed from left to right and top downwards in the apertures in the rows of the indicated number of columns and the plain text then reappears in the apertures in the columns, reading downward and from left to right. (It is, of course, not essential that nulls be added in encipherment to make the length of the cryptogram an exact multiple of 5, for the matter can readily be handled even if this is not done. In decipherment the total number of letters divided by 5 will give the number of complete columns; the remainder left over from the division will give the number of cells occupied by letters in the last column on the right.)

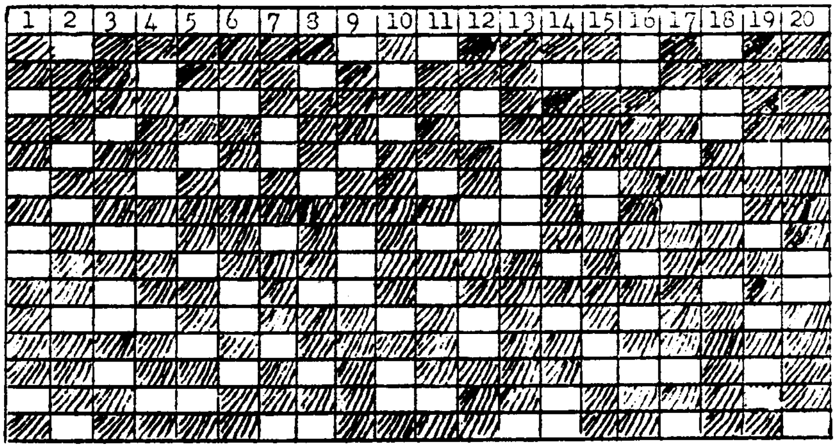


Figure 46.

b. Such a grille can assume 4 positions, two obverse and two reverse. Arrangements must be made in advance as to the sequence in which the various positions will be employed.

c. The solution of a single cryptogram enciphered by one and only one position of such a grille presents a practically hopeless problem, for the apertures being distributed at random throughout the grille there is nothing which may be seized upon as a guide to the reconstruction of either the grille or the plain text. It is conceivable, of course, that a person with an infinite amount of patience could produce an intelligible text and a grille conformable to that text, the grille having a definite number of columns and a fixed number of apertures distributed at random throughout the columns. But there would be no way of proving that the plain text so obtained is the actual plain text that was enciphered; for it would be possible to produce several "solutions" of the same character, any one of which might be correct.[4]

d. However, suppose a grille of this sort were employed to encipher a long message, requiring two or more applications of the grille. For example, in the case of the grille shown in figure 46, having a capacity of 100 letters per application, suppose a message of 400 letters were to be enciphered, requiring two obverse and two reverse applications of the grille. It is obvious that symmetrical relationships of the nature of those pointed out in paragraph 31 can be established. Of course, if the grille is used several times in the same position to its full capacity, producing cryptograms of multiples of 100 letters, then the sections of 100 letters may be superimposed and the general solution elucidated in paragraph 26 applied.

---

[4] In this connection, see *Military Cryptanalysis, Part III*, sec. XI, footnote 8.

e. If the grille shown in figure 46 were used to encipher two messages, one of 80 letters, the other of 85, it would be possible to solve these messages. For by eliminating 5 letters from the longer message, the two cryptograms can be superimposed and handled as in paragraph 26. The difficulty would be in finding the 5 extra letters. Of course, if it should happen that one of the messages required 3 or 4 nulls and letters such as J, X, or Z were employed for this purpose, the nulls would be likely characters for elimination. But regardless of this, even if letters of medium or high frequency were used as nulls, patient experimentation would ultimately lead to solution. The latter, it must be conceded, would be difficult but not impossible.

SECTION VIII

COMBINED SUBSTITUTION-TRANSPOSITION SYSTEMS

Paragraph
Reasons for combining transposition with substitution 35
Other types of combined substitution-transposition systems 36

35. Reasons for combining transposition with substitution.—*a.* Transposition methods are, from the cryptographic point of view, rather highly regarded because they are, as "hand methods" go, rather rapid in operation and usually quite simple. However, from their very nature they entail the disadvantage that a single-letter omission or addition may render their decryptographing difficult if not impossible for the average cryptographic clerk. But from the standpoint of modern cryptography the principal disadvantage of transposition methods is that they can be mechanized only with great difficulty—certainly with greater difficulty than is the case of substitution methods. Only one or two attempts have been made to produce machinery for effecting transposition, and these have not been successful.

b. Pure transposition, that is, transposition by itself, without an accompanying substitution or other means of disguise for the letters of the plain text, hardly affords sufficient guarantees for cryptographic security in the case of a voluminous correspondence which must be kept really secret for any length of time. For no matter how complex the method, or how many transpositions may be applied to the letters of a single message, sight must never be lost of the fact that when there are many messages in the same key there are bound to be two or more of identical length; and when this is the case the type of solution described in paragraph 26 may be applied to these cryptograms, the transposition keys recovered, and then all other messages in the same key translated.

c. A message may undergo monoalphabetic substitution and the resulting text passed through a simple transposition. When this is the case a uniliteral frequency distribution will, of course, exhibit all the characteristics of monoalphabeticity, yet the cryptogram will resist all attempts at solution according to straightforward simple substitution principles. It is usually not difficult to detect that a transposition is involved because there will not only be long strings of low-frequency letters or high-frequency letters but what is more important, *there will be very few or no repetitions of digraphs, trigraphs, and tetragraphs, since these will be broken up by the transposition.* When a uniliteral distribution presents all the external evidences of monoalphabeticity and yet there are no repetitions, it is almost a positive indication of the presence of transposition superimposed upon the substitution, or vice versa. (The former is usually the case.)

d. When confronted with such a situation the cryptanalyst usually proceeds by stages, first eliminating the transposition and then solving the substitution. It is of course obvious that the general solution for transposition ciphers (cryptograms of identical length in the same key) will not be applicable here, for the reason that such a solution is based upon anagramming, which in turn is guided by the disclosure of good digraphs, trigraphs, and polygraphs. Since the letters of a combined substitution-transposition cipher are no longer the same as the original plain-text letters, simple anagramming of columns formed by superimposing identical-length cryptograms can yield no results, because there is nothing of the nature of plain text to guide the cryptanalyst in his juxtaposition of columns.[1]

e. Of course, if it should happen that the substitution process involves known alphabets, the cryptanalyst can remove the effects of the substitutive process before proceeding to eliminate the transposition, even if in the encipherment the substitution came first. For example, if a standard cipher alphabet were employed for the substitution the uniliteral frequency distribution would give indications thereof and the cipher letters could immediately be converted to the normal plain-text equivalents. The latter may then be studied as though merely transposition had been applied. But if unknown mixed cipher alphabets were employed, this initial step can not be accomplished and a solution must usually wait upon the removal of the transposition before the substitution can be attacked. The latter may be very difficult or impossible where a good transposition method is used; where simple columnar transposition is used the removal of the transposition can be effected if the message is long enough.

f. Of course if nothing is known about the system of transposition that has been employed, there is hardly anything to do but experiment with various types of transposition in an attempt to bring about such an arrangement of the text as will show repetitions. If this can be done, then the problem can be solved. For example, suppose that a message has been enciphered by a single mixed cipher alphabet and the substitution text has then been inscribed within a rectangle of certain dimensions according to one of the usual routes mentioned in paragraph 5 of this text. Repetitions in the plain text will of course be preserved in the substitution text but will be destroyed after the transposition has been applied. The cryptanalyst, however, in his attempts to eliminate the transposition, may experiment with route transpositions of the various types, employing rectangles of various dimensions as suggested by the total number of letters in the cryptogram. If he perseveres, he will find one route which he will know is correct as soon as he tries it *because it will disclose the repetitions in the plain text,* although the latter are still covered by a substitution.

g. Practically all the methods of transposition which may be applied to plain text may also be applied to a text resulting from an initial transformation by substitution. As already mentioned, route transposition may be used; reversed and rail-fence writing, columnar transposition with or without keying and with complete or incomplete rectangles are also possible. From a practical standpoint, keyed-columnar transposition applied to a monoalphabetic substitution is not only a popular but also a fairly secure combination because in this case the elimination of the transposition is a rather difficult matter. If the rectangle is completely filled the problem is not insurmountable in the case of a long message transposed by means of transposition with a rectangle of fairly small dimensions. For by assuming rectangles of various dimensions suggested by the total number of letters, cutting the columns apart, and then combining columns on the basis of the number of repetitions produced within juxtaposed columns and between different sets of juxtaposed columns, it is possible to reconstruct the rectangle and thus remove the transposition phase. This, however, is admittedly a slow and difficult process even under the most favorable conditions; and if the rectangle is incompletely filled the process is very difficult. For in the latter case the lack of absolutely clear-cut knowledge as to the lengths of the columns, the juxtaposition of columnar material becomes replete with uncertainties and engenders feelings of confusion, hopelessness, and inadequacy in the mind of the cryptanalyst. However, he need not be wholly in despair if he is confronted with a problem of this nature in war time, when many cryptograms become available for study. For there are special methods of solution suitable to the occasion, created by special circumstances attendant upon the interception of a voluminous traffic. In subsequent paragraphs the student will come to understand what is here meant by the special circumstances and will learn of these special solutions.

---

[1] It should, however, not be inferred that anagramming processes are entirely excluded in the cryptanalysis of all combined substitution-transposition systems. In certain cases the anagramming process may be guided by considerations of frequency of letters or fragments of letters. A case of this kind will be encountered in the solution of the ADFGVX cipher, par. 40.

36. Other types of combined substitution-transposition systems.—*a.* There is no technical obstacle to the application of a transposition to the text resulting from any type of substitution, even if the latter is polyalphabetic or polygraphic in nature. The obstacles, or rather objections, to such combinations are practical in their character—they are too complex for ordinary use and the prevalence of errors makes them too difficult to handle, as a general rule. However, they have been and are sometimes used even as field ciphers. For instance, on the southeastern front during the World War, the Central Powers made use of a somewhat irregular polyalphabetic substitution involving four standard alphabets and a keyed columnar transposition with incompletely-filled rectangles of a relatively large number of columns. Nevertheless, messages in this system were solved by taking advantage of the possibility of devising special solutions.

b. A few remarks may be made in regard to the order in which the two processes, substitution and transposition, are employed in a combined system. It is clear that when the substitution is monoalphabetic it is immaterial, so far as cryptographic security is concerned, whether substitution is followed by transposition or vice versa, because the equivalent of each plain-text letter remains fixed regardless of the order in which the plain-text letters appear in the plain text. However, if the substitution is polyalphabetic in character it is better that the transposition process precede the substitution process, and that the number of alphabets employed be different from the number of elements in the transposition key, if columnar transposition is the case. The best situation, from a cryptographic security standpoint, is when the two key lengths (substitution and transposition) have no common factor. If the two keys are of the same length, the letters in each column are enciphered by the same cipher alphabet and thus the cryptogram would contain a certain number of sections of approximately equal length, composed of letters falling in the same cipher alphabet.

c. Digraphic substitution, such as that produced by the Playfair Cipher, may be combined with transposition to yield cryptograms of fair security. But here again the elimination of the transposition phase by taking advantage of special circumstances or by rearranging the next so as to uncover the repetitions which are inevitable in the Playfair Cipher, will result in solution.

d. A particularly fruitful source of combined substitution-transposition is to be found in those methods generally designated as fractionating systems, in which in the substitution phase each plain-text letter is replaced by an equivalent composed of two or more components or "fractions" and then these components are subjected to transposition in a second phase. This latter may be followed by a third phase, recombination of distributed components, and a fourth phase, the replacement of the recombined components by letters. Thus, such a system comprises a first substitution, a transposition, a recombination, and a second substitution.[2] In the subsequent paragraphs certain systems of this sort will be dealt with in detail. They are interesting examples of practical systems of cryptography which have been used in the field of military operations in the past and may again be used in the future. The first one to be discussed is particularly interesting for this reason alone; but it is also of interest because it will serve as a model for the student to follow in his study of methods for the solution of combined substitution-transposition ciphers in general.

---

[2] See Special Text No. 166, *Advanced Military Cryptography*, sec. XI.

SECTION IX

SOLUTION OF THE ADFGVX SYSTEM

Paragraph

Introductory remarks 37

Special solution by means of identical endings 38

Special solution by means of identical beginnings 39

Special solution by the exact factor method 40

General solution for the ADFGVX system 41

Basic principles of the general solution 42

Illustration of solution 43

37. Introductory remarks.—*a.* One of the most interesting and practical of the many methods in which substitution and transposition are combined within a single system is that known in the literature as the ADFGVX cipher.[1] In this system a 36-character bipartite substitution checkerboard is employed, in the cells of which the 26 letters of the alphabet and the 10 digits are distributed in mixed order, often according to some key word. The row and column indicators (coordinates) are the letters ADFGVX, and, taken in pairs, the latter are used as substitutes for the letters of the plain text. These substitutive pairs are then inscribed within a rectangle and a columnar transposition takes place, according to a numerical key. The cipher text consists then merely of the 6 letters A, D, F, G, V, and X.

b. The ADFGVX cipher system was inaugurated on the Western Front by the German Army on March 1, 1918, for communication between higher headquarters, principally between headquarters of divisions and corps. When first instituted on March 1, 1918, the checkerboard consisted of 25 cells, for a 25-letter German alphabet (J was omitted), and the 5 letters A, D, F, G, and X used as coordinates. On June 1 the letter V was added, the checkerboard having been enlarged to 36 cells, to take care of a 26-letter alphabet plus the 10 digits. Transposition keys ranged from 15 to 22 numbers, inclusive, and both the checkerboard and the transposition key were changed daily. The number of messages in this system varied from 25 a day upon the inception of the system to as many as 150 per day, during the last days of May 1918. The first solution was made on April 6 by the French. The cipher continued in use rather extensively until late in June but from that time until the Armistice the volume of messages diminished very considerably. Although only 10 keys, covering a period of as many days were ever solved, the proportion of solved messages in the whole intercepted traffic was about 50 percent. This was true because of the fact that the keys solved were those for days on which the greatest number of messages was intercepted. The same system was employed on the southeastern front from July 1918 to the end of the war. Keys were in effect at first for a period of 2 days and beginning on September 1, for a period of 3 days. In all 17 keys, covering a total of 44 days, were solved.

c. At the time that the Allied cryptanalytic offices were working with cryptograms in this system only three methods were known for their solution and all three of them are classifiable under the heading of *special solutions*, because certain conditions had to obtain before they could be applied. No general solution had been developed until after hostilities had ceased.[2]

---

[1] Special Text No. 166, *Advanced Military Cryptography*, sec. XI.

[2] The general solution to be described in paragraphs 41–43 was not established until after the Armistice. Had it been elaborated earlier there would no doubt have been many more solutions than were actually effected by the methods then available.

Because they are interesting and useful some attention will be devoted to both the general and the special solutions. Since the special solutions are easy to understand and serve as a good introduction to the general solution, they will be taken up first.

38. Special solution by means of identical endings.—*a*. In paragraph 24 it was demonstrated how the solution of keyed-columnar transposition ciphers can be facilitated and simplified by the comparison of two cryptograms which are in the same key and the plain-text endings of which are identical. It was noted in that case that a study of the irregularly distributed cipher-text identities between the two cryptograms permits of not only cutting up the text into sections that correspond with the long and the short columns of the transposition rectangle but also of establishing the transposition key in a direct manner almost entirely mathematical in nature. When this has been accomplished the plain texts of these two messages are at once disclosed, and all other messages in the same key may be read by means of the key so reconstructed.

b. The same method of solution is applicable to the similar situation, if it can be found, in the case of the ADFGVX system, except that one more step intervenes between the reconstruction of the transposition rectangle and the appearance of the plain text in the rectangle: A monoalphabetic substitution must be solved, since the text in the rows of the rectangle does not consist of plain-text letters but of pairs of components representing these letters as enciphered by means of a bipartite substitution alphabet. Moreover, this latter step is comparatively simple when there is a sufficient amount of text in the two rectangles; if not, additional material for use in solving the monoalphabet can be obtained from other cryptograms in the same key, if they are available, since the transposition key, having already been reconstructed from the two cryptograms with identical endings, will permit of inscribing all other cryptograms in the same key within their proper rectangles.

c. A demonstration of the application of the principles involved in such a solution will be useful. The following cryptograms have been intercepted on the same date, the 20th:

No. 1

To CG 22d Brigade:

```
        5         10        15        20        25        30
X V A A X V D D A G D A D V F A D A D A F X G F V X F A X A
        35        40        45        50        55        60
X V A V F A V X A D G F F X F F G A G F D G D G D D G A F D
        65        70        75        80        85        90
A A D D D X D A V G G A A D X A D F V F F D F X F G F G A V
        95        100       105       110       115       120
A F A F X F F X F X F V D V X A F F G X A A A V A V A F A G
        125       130       135       140       145       150
D D F A G V F A D V F A V V X G V A A A F D F A X X F A A G

D X
```

No. 2

To CG 23d Brigade:

```
        5          10          15          20          25          30
F D F F F   F V F A D   D V F V D   G A F D F   D A G A D   F D F A F
       35          40          45          50          55          60
V A X G D   V X G F X   V X D X V   A A A A D   G X F F D   V F A A G
       65          70          75          80          85          90
V G V F F   F D A F F   F X D A F   X G A F D   V F V X V   D D F A D
       95         100         105         110         115         120
D A A A X   A A F F A   F V F X F   F A X X A   A D V X A   V D A V F
      125         130         135         140         145         150
D F A V X   V A D X F   A X F F X   X A A V X   X A D X A   A A V V G
      155         160         165         170         175         180
A G D X X   F D F A X   F D G D F   F X D G X   F A G D F   F D D V D
      185         190
D X D A F   A G X X A   F G A V
```

d. The delimitation and marking of identities between these two cryptograms is a procedure similar to that explained in paragraph 24*b*, except that a little more study may be necessary in this case because occasionally there may be considerable uncertainty as to exactly where an identity begins or ends. The reason for this is not difficult to understand. Whereas in paragraph 24*b* the process involves "unfractionated" letters and there are about 18 or 20 different letters to deal with, so that an "accidental identity" is a rather rare occurrence, in the present problem the process involves fractions of letters (the components of the bipartite cipher equivalents), and there are only 6 different characters to deal with, so that such "accidental identities" are quite frequent. Now the cryptanalyst is not able at first to distinguish between these accidental identities and actual identities and this is what makes the process somewhat difficult. What is meant will become perfectly clear presently.

e. Taking the two illustrative cryptograms, the first step is to ascertain what identities can be found between them, and then mark off these identities. For example, it is obvious that if the messages end alike the last several letters in No. 1 should be found somewhere in No. 2, and likewise the last several letters in No. 2 should be found somewhere in No. 1. The number of letters in identical sequences will depend upon the length of the identical text and the width of the transposition rectangle. Searching through No. 2 for a sequence such as A G D X, or G D X, or at least D X, the tetragraph A G D X is found as letters 151–54. The last column of No. 2 ends with F G A V; searching through No. 1 for a sequence F G A V, or G A V, or at least A V, the tetragraph F G A V is found as letters 87–90. These identities are underlined or marked off in some fashion, and search is made for other identities. It would be a great help if the width of the transposition rectangle were known, for then it would be possible to cut up the text into lengths approximately corresponding to column lengths, and this would then restrict the search for identical sequences to those sections which correspond to the bottoms of the columns. Suppose the key to contain 20 numbers. Then the rectangle for No. 1, containing 152 letters, would consist of 12 long columns of 8 letters and 8 short ones of 7 letters; that for No. 2, containing 194 letters, would consist of 14 long columns of 10 letters and 6 short ones of 9 letters. If that were correct then in No. 1 the end of the first column would be either X V D D, or X V D. Searching through No. 2 for either of these a sequence X V D D is found as letters 84–7. Column 1 is probably a long column in No. 1. The word *probably* is used because the identity may extend only over the letters X V D, *and the next* D *may be an accidental similarity*, since the chances that D will appear by pure accident are 1 in 6, which is not at all improbable. It must also be pointed out that a certain number of telegraphic errors may be expected, and since there are only 6 different letters the chances that an F, for example, will be received or recorded as a D are fairly good. Column 1 of No. 2 ends either with V F A D or V F A. Searching through No. 1, a sequence V F A D is found as letters 14–17; a sequence V F A is found as letters 34–6; a sequence V F F D is found as letters 79–82; a sequence V F A D is also found as letters 126–130; a sequence V F A is found as letters 130–2. Here are several possibilities; which is the one to choose? Two of these possibilities coincide exactly with the full sequence being sought, V F A D. Can one of them be eliminated as a possibility? Perhaps tables to facilitate the location of possible "breaks" will be helpful in making the elimination (see paragraph 16*n*). "Break tables" are therefore constructed for the messages on the basis of rectangles of 20 columns, and are as shown below.

	0	8	16	24	32	40	48	56	64	72	80	88	96
0	0	8	16	24	32	40	48	56	64	72	80	88	96
7	7	15	23	31	39	47	55	63	71	79	87	95	103
14	14	22	30	38	46	54	62	70	78	86	94	102	110
21	21	29	37	45	53	61	69	77	85	93	101	109	117
28	28	36	44	52	60	68	76	84	92	100	108	116	124
35	35	43	51	59	67	75	83	91	99	107	115	123	131
42	42	50	58	66	74	82	90	98	106	114	122	130	138
49	49	57	65	73	81	89	97	105	113	121	129	137	145
56	56	64	72	80	88	96	104	112	120	128	136	144	152

"Break" table for No. 1 (152 letters)

	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140
0	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140
9	9	19	29	39	49	59	69	79	89	99	109	119	129	139	149
18	18	28	38	48	58	68	78	88	98	108	118	128	138	148	158
27	27	37	47	57	67	77	87	97	107	117	127	137	147	157	167
36	36	46	56	66	76	86	96	106	116	126	136	146	156	166	176
45	45	55	65	75	85	95	105	115	125	135	145	155	165	175	185
54	54	64	74	84	94	104	114	124	134	144	154	164	174	184	194

"Break" table for No. 2 (194 letters)

From these tables it follows that as regards message No. 1 there can be a break after the 7th, 8th, 14th, 15th, 16th . . . letters but not after the 6th letter, nor after the 9th to 14th letters, nor after the 17th to 21st letters, and so on; as regards message No. 2 there can be a break after the 9th, 10th, 18th, 19th, 20th, . . . letters but not after the 8th letter nor after the 11th to 18th letters, nor after the 21st to 27th letters, and so on. Referring again to the two **VFAD** sequences in No. 1 which may correspond with the **VFAD** sequence in No. 2, it was found that the first candidate would require a break immediately after the 17th letter. But the break table for No. 1 precludes this possibility; hence the first **VFAD** sequence in No. 1 in position 14–17 may be eliminated as a candidate, leaving the second **VFAD**, in position 126–130, as a candidate. This would require a break after the 130th letter and reference to the break table for No. 1 shows this to be a possibility. Hence, the **VFAD** in position 126–130 in No. 1 will tentatively be accepted as matching the **VFAD** sequence in No. 2. Another section of the text of one or the other cryptogram is next selected, with a view to establishing additional identities. To go through the whole process here would consume too much space and time. Moreover, it is not necessary, for the only purpose in carrying the demonstration this far is to indicate to the student the general procedure and to show him some of the difficulties he will encounter in the identification of the similar portions when the text is composed of only a very limited number of different letters. In this case, after more or less tedious experimentation, the hypothesis of a key of 20 columns is established as correct, whereupon two sets of 20 identities are uncovered and the identities are found to be as shown below.

No. 1

$$\mathrm{X\;V\;A\;A}\;\underbrace{\overset{5}{\mathrm{X}}\;\;\mathrm{V\;D\;D}}_{1}\;\mathrm{A}\;\overset{10}{\mathrm{G}}\quad \mathrm{D\;A}\;\underbrace{\mathrm{D\;V}\;\overset{15}{\mathrm{F}}}_{2}\quad \mathrm{A\;D\;A}\;\underbrace{\mathrm{D}\;\overset{20}{\mathrm{A}}\;\;\mathrm{F\;X\;G}}_{3}\;\mathrm{F}\;\overset{25}{\mathrm{V}}\quad \mathrm{X\;F}\;\underbrace{\mathrm{A\;X}\;\overset{30}{\mathrm{A}}}_{4}$$

$$\mathrm{X\;V\;A\;V}\;\underbrace{\overset{35}{\mathrm{F}}\;\;\mathrm{A\;V\;X}}_{5}\;\mathrm{A}\;\overset{40}{\mathrm{D}}\quad \mathrm{G\;F}\;\underbrace{\mathrm{F\;X}\;\overset{45}{\mathrm{F}}}_{6}\quad \mathrm{F\;G\;A\;G}\;\underbrace{\overset{50}{\mathrm{F}}\;\;\mathrm{D\;G\;D}}_{7}\;\mathrm{G}\;\overset{55}{\mathrm{D}}\quad \underbrace{\mathrm{D\;G\;A\;F}\;\overset{60}{\mathrm{D}}}_{8}$$

$$\mathrm{A\;A\;D}\;\underbrace{\mathrm{D}\;\overset{65}{\mathrm{D}}\;\;\mathrm{X\;D\;A}}_{9}\;\mathrm{V}\;\overset{70}{\mathrm{G}}\quad \mathrm{G\;A}\;\underbrace{\mathrm{A\;D}\;\overset{75}{\mathrm{X}}}_{10}\quad \mathrm{A\;D\;F\;V}\;\underbrace{\overset{80}{\mathrm{F}}\;\;\mathrm{F\;D}}_{11}\;\mathrm{F\;X}\;\overset{85}{\mathrm{F}}\quad \mathrm{G}\;\underbrace{\mathrm{F\;G\;A}\;\overset{90}{\mathrm{V}}}_{12}$$

$$\mathrm{A\;F\;A\;F}\;\underbrace{\overset{95}{\mathrm{X}}\;\;\mathrm{F\;F\;X}}_{13}\;\mathrm{F}\;\overset{100}{\mathrm{X}}\quad \mathrm{F\;V}\;\underbrace{\mathrm{D\;V}\;\overset{105}{\mathrm{X}}\;\;\mathrm{A}}_{14}\;\mathrm{F\;F\;G}\;\overset{110}{\mathrm{X}}\quad \underbrace{\mathrm{A\;A\;A}}_{15}\;\mathrm{V}\;\overset{115}{\mathrm{A}}\quad \mathrm{V\;A}\;\underbrace{\mathrm{F\;A}\;\overset{120}{\mathrm{G}}}_{16}$$

$$\underline{\mathrm{D}}\;\mathrm{D\;F\;A}\;\overset{125}{\mathrm{G}}\quad \underbrace{\mathrm{V\;F\;A\;D}}_{17}\;\overset{130}{\mathrm{V}}\quad \mathrm{F\;A\;V}\;\underbrace{\mathrm{V}\;\overset{135}{\mathrm{X}}\;\;\mathrm{G}}_{18}\;\mathrm{V\;A\;A}\;\overset{140}{\mathrm{A}}\quad \underbrace{\mathrm{F\;D\;F\;A}}_{19}\;\overset{145}{\mathrm{X}}\quad \mathrm{X\;F\;A}\;\underbrace{\mathrm{A}\;\overset{150}{\mathrm{G}}}$$

$$\underbrace{\mathrm{D\;X}}_{20}$$

FIGURE 47.

No. 2

$$\text{F D F F }\overset{5}{\text{F}}\quad \text{F }\underbrace{\text{V F A }\overset{10}{\text{D}}}_{1}\quad \text{D V F V }\underbrace{\overset{15}{\text{D}}\ \text{G A F D}}_{2}\,\overset{20}{\text{F}}\quad \text{D A G A }\overset{25}{\text{D}}\quad \underbrace{\text{F D F A}}_{3}\,\overset{30}{\text{F}}$$

$$\text{V A X G }\overset{35}{\text{D}}\quad \underbrace{\text{V X G}}_{4}\text{ F }\overset{40}{\text{X}}\quad \text{V X D X }\overset{45}{\text{V}}\quad \underbrace{\text{A A}}_{5}\text{ A A }\overset{50}{\text{D}}\quad \text{G X F F }\underbrace{\overset{55}{\text{D}}\ \text{V F}}_{6}\text{ A A }\overset{60}{\text{G}}$$

$$\text{V G V F }\underbrace{\overset{65}{\text{F}}\ \text{F D}}_{7}\text{ A F }\overset{70}{\text{F}}\quad \text{F X }\underbrace{\text{D A }\overset{75}{\text{F}}\ \text{X G}}_{8}\text{ A F }\overset{80}{\text{D}}\quad \text{V F V }\underbrace{\text{X }\overset{85}{\text{V}}\ \text{D D}}_{9}\text{ F A }\overset{90}{\text{D}}$$

$$\text{D A A }\underbrace{\text{A }\overset{95}{\text{X}}\ \text{A}}_{10}\text{ A F F }\overset{100}{\text{A}}\quad \text{F V }\underbrace{\text{F X }\overset{105}{\text{F}}}_{11}\quad \text{F A X X }\overset{110}{\text{A}}\quad \text{A }\underbrace{\text{D V X }\overset{115}{\text{A}}}_{12}\quad \text{V D A V }\overset{120}{\text{F}}$$

$$\text{D }\underbrace{\text{F A V }\overset{125}{\text{X}}}_{13}\quad \text{V A D X }\overset{130}{\text{F}}\quad \text{A }\underbrace{\text{X F F }\overset{135}{\text{X}}}_{14}\quad \text{X A A V }\overset{140}{\text{X}}\quad \text{X }\underbrace{\text{A D X}}_{15}\,\overset{145}{\text{A}}\quad \text{A A V V }\overset{150}{\text{G}}$$

$$\underbrace{\text{A G D X}}_{16}\,\overset{155}{\text{X}}\quad \text{F D F A }\overset{160}{\text{X}}\quad \underbrace{\text{F D G D}}_{17}\,\overset{165}{\text{F}}\quad \text{F X D G }\overset{170}{\text{X}}\quad \underbrace{\text{F A G D}}_{18}\,\overset{175}{\text{F}}\quad \text{F D D V }\underbrace{\overset{180}{\text{D}}}$$

$$\underbrace{\text{D X D A}}_{19}\,\overset{185}{\text{F}}\quad \text{A G X X }\overset{190}{\text{A}}\quad \underbrace{\text{F G A V}}_{20}$$

FIGURE 47—Continued

f. A table of equivalencies[3] is then drawn up:

No. 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. 2	9	6	8	10	13	11	17	2	19	15	7	20	14	12	5	18	1	4	3	16

Since the rectangle for No. 2 has 2 more letters in the last row than the rectangle for No. 1, two chains of equivalents at two intervals are constructed. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		9		19		3		8		2		6		11		7		17	
4		10		15		5		13		14		12		20		16		18	

These chains must now be united into a single chain by proper interlocking. Since cryptogram No. 1 has 12 long columns, and since the identities of these 12 columns are now known (1, 3, 5, 7, 9, 12, 13, 14, 16, 17, 19, 20), the interlocking of the two chains and hence the transposition key must be this:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
7	5	17	13	1	14	9	12	19	20	3	16	8	18	2	4	6	10	11	15

g. The two cryptograms may now be transcribed into their proper transposition matrices, as shown in figure 48.

---

[3] It is necessary to remark that in setting up the table of equivalencies, after determining the width of the rectangle, that message which has the lesser number of long columns is used as the basis for the normal sequence 1, 2, 3, ... If the one having the greater number of long columns is employed as the base, the reconstructed key will be reversed.

No. 1

7	5	17	13	1	14	9	12	19	20	3	16	8	18	2	4	6	10	11	15
F	X	D	A	X	F	A	F	V	X	A	V	G	V	A	F	A	V	A	F
G	V	F	F	V	X	A	X	A	X	D	A	D	F	G	V	D	G	D	F
A	A	A	A	A	F	D	F	A	F	A	V	D	A	D	X	G	G	F	G
G	V	G	F	A	V	D	G	A	A	D	A	G	V	A	F	F	A	V	X
F	F	V	X	X	D	D	F	F	A	A	F	A	V	D	A	F	A	F	A
D	A	F	F	V	V	X	G	D	G	F	A	F	X	V	X	X	D	F	A
G	V	A	F	D	X	D	A	F	D	X	G	D	G	F	A	F	X	D	A
D	X	D	X	D	A	A	V	A	X	G	D								

No. 2

7	5	17	13	1	14	9	12	19	20	3	16	8	18	2	4	6	10	11	15
A	F	X	V	F	V	A	F	F	F	F	A	A	F	D	F	A	F	A	X
A	X	F	D	D	A	F	A	F	A	D	A	F	F	V	V	D	A	F	A
G	V	D	A	F	D	D	X	D	G	A	A	F	X	F	A	G	D	F	A
V	X	F	V	F	X	V	X	D	X	G	V	F	D	V	X	X	D	A	V
G	D	A	F	F	F	F	A	V	X	A	V	X	G	D	G	F	A	F	X
V	X	X	D	F	A	V	A	D	A	D	G	D	X	G	D	F	A	V	X
F	V	F	F	V	X	X	D	D	F	F	A	A	F	A	V	D	A	F	A
F	A	D	A	F	F	V	V	X	G	D	G	F	A	F	X	V	X	X	D
F	A	G	V	A	F	D	X	D	A	F	D	X	G	D	G	F	A	F	X
D	A	D	X	D	X	D	A	A	V	A	X	G	D						

FIGURE 48.

h. A frequency distribution is now made of all the bipartite pairs, so as to solve the enciphering checkerboard. There is no necessity for going through this part of the solution, for it falls along quite normal lines of monoalphabetic substitution. The two plain-text rectangles are shown in figure 49. The checkerboard [4] is found to be as shown in figure 50*a*.

No. 1

7 5	17 13	1 14	9 12	19 20	3 16	8 18	2 4	6 10	11 15
H	A	V	E	O	R	D	E	R	E
F X	D A	X F	A F	V X	A V	G V	A F	A V	A F
D	C	O	M	M	A	N	D	I	N
G V	F F	V X	A X	A X	D A	D F	G V	D G	D F
G	G	E	N	E	R	A	L	2	3
A A	A A	A F	D F	A F	A V	D A	D X	G G	F G
D	B	R	I	G	A	D	E	T	O
G V	G F	A V	D G	A A	D A	G V	A F	F A	V X
C	O	U	N	T	E	R	A	T	T
F F	V X	X D	D F	F A	A F	A V	D A	F A	F A
A	C	K	W	I	T	H	O	U	T
D A	F F	V V	X G	D G	F A	F X	V X	X D	F A
D	E	L	A	Y	W	I	T	H	A
G V	A F	D X	D A	F D	X G	D G	F A	F X	D A
L	L	A	R	M	S				
D X	D X	D A	A V	A X	G D				

FIGURE 49.

---

[4] Since the second cryptogram is addressed to the CG 23d Brigade and the first cryptogram mentions that the commander of that brigade has been ordered to do so and so, the solution of the groups GG (=2) and FG (=3) is made by inference. This gives the placement of these two digits in the cipher square.

No. 2

7 5	17 13	1 14	9 12	19 20	3 16	8 18	2 4	6 10	11 15
E	X	P	E	C	T	E	N	E	M
A F	X V	F V	A F	F F	F A	A F	D F	A F	A X
M	Y	A	T	T	A	C	K	A	T
A X	F D	D A	F A	F A	D A	F F	V V	D A	F A
D	A	Y	L	I	G	H	T	S	T
G V	D A	F D	D X	D G	A A	F X	F A	G D	F A
O	P	H	O	L	D	Y	O	U	R
V X	F V	F X	V X	D X	G V	F D	V X	X D	A V
S	E	C	T	O	R	W	I	T	H
G D	A F	F F	F A	V X	A V	X G	D G	F A	F X
O	U	T	F	A	I	L	S	T	O
V X	X D	F A	V A	D A	D G	D X	G D	F A	V X
P	C	O	U	N	T	E	R	A	T
F V	F F	V X	X D	D F	F A	A F	A V	D A	F A
T	A	C	K	W	I	T	H	O	U
F A	D A	F F	V V	X G	D G	F A	F X	V X	X D
T	D	E	L	A	Y	W	I	T	H
F A	G V	A F	D X	D A	F D	X G	D G	F A	F X
A	L	L	A	R	M	S			
D A	D X	D X	D A	A V	A X	G D			

FIGURE 49—Continued.

2nd component

1st component	A	D	F	G	V	X
A	G		E		R	M
D	A		N	I		L
F	T	Y	C	3	P	H
G		S	B	2	D	F
V				K		0
X		U	V	W	X	

FIGURE 50a.

	A	D	F	G	V	X
A	G		E		R	M
D	A		N	I		L
F	T	Y	C	3	P	H
G		S	B	2	D	
V	F				K	0
X	(Q)	U	V	W	X	(Z)

FIGURE 50b.

	A	D	F	G	V	X
A	G	6	E	5	R	M
D	A	1	N	I	8	L
F	T	Y	C	3	P	H
G	7	S	B	2	D	4
V	F	6	J	0	K	0
X	Q	U	V	W	X	Z

FIGURE 50c.

i. Speculating upon the disposition of the letters within the enciphering checkerboard, it becomes evident that the key phrase upon which it is based is GERMAN MILITARY CIPHERS. The fact that the digit 2 follows B and the digit 3 follows C suggests that the digits are inserted immediately after the letters A, B, C, . . ., as they occur in the mixed sequence. Note the cells which still remain vacant after the key word mixed sequence is fully developed in the checkerboard, and all the letters which do occur in the two messages are inserted in their correct cells (fig. 50*b*). The complete checkerboard may therefore be taken almost certainly to be as shown in figure 50*c*. The date (20th) indicates that the transposition key will have 20 numbers in it. The transposition key was evidently derived from the first 20 letters of the mixed sequence:

G	E	R	M	A	N	I	L	T	Y	C	P	H	S	B	D	F	J	K	O
7	5	17	13	1	14	9	12	19	20	3	16	8	18	2	4	6	10	11	15

39. Special solution by means of identical beginnings.—*a*. In paragraph 23 was demonstrated the method of solution based upon finding two cryptograms which are in the same key and the plain texts of which begin with the same words. The application of this method to the corresponding situation in the case of the ADFGVX system should by this time be obvious. The finding of identical sequences is somewhat easier in this case than in the case of identical endings because the identities can be found in parallel progression from the beginning to the end of the two cryptograms being compared. Moreover, the discovery of two cryptograms with similar beginnings is easier than that of two with similar endings because in the former case the very first groups in the two cryptograms contain identities, whereas in the latter case the identities are hidden and scattered throughout the texts of the two cryptograms. On the other hand, the complete solution of a case of identical endings is very much more simple than that involving identical beginnings because in the former case the establishment of the identities carries with it almost automatically the complete reconstruction of the transposition key, whereas in the latter this is far from true and additional cryptograms may be essential in order to accomplish this *sine qua non* for the solution.

b. The following represent 8 cryptograms of the same date, assumed to have been enciphered by the same key.

No. 1

```
VDDFA XFAAX DXGGF FVFXF GXDXG DGAGF
AGDAD VGGDA AADXX DXAFF AADAF DFFDA
```

No. 2

```
GXDDA DDGDF VGXAX XXGXG AAAAD FADDX
AVDXF XAD
```

No. 3

```
XDAAA GXDDX VFFVD GADFD XAAAG DFADG
AFDAD GVGDV FDFXA GFXAF AFAXD DDDFD
XAXVA DXFXF DGAGF GGADD AGDGX AVGDG
ADAFA XFAAG VAAGA FDVDV DXFDA XFDFF
GDXDV DADAV DADDD GADAG AAAFG GDXAX
FGVXD DGDDF AFAGV AFGXG VDDAX XDVFF
FFDXG VGDFG AVADA XDAFA AFDGF VFXXX
AAGAG AFDGX AFAFX XGGAG AAFFA AFDGA
GAFVX DGGFG DAAAF DADAD XVVAX FVADD
GAFFF GXAXD FDDFX AAAAA
```

No. 4

```
AFGFX AGXAG XDDAF AAXAV GDDDD FAFGV
DGDXA FDXAX GFGDD VADXA XGFAX FDADD
GD
```

No. 5

```
XAAAD DGAAG DDDXF FAVGA XDGGD FFAVA
DAAXA GDXDX XXXDG VFADA DFFFF VVGFD
XFDGG DAXDG ADFD
```

No. 6

```
X D A A V   D X D G F   X V G D D   A V G X A   D X A A D   X G G A A
G D F D A   A A G A X   D V F D F   D F F D D   F D D F X   F X X F D
F D X A X   G A X F F   V D V A F   G V D V D   D D A G D   G G D A A
G G F D D   D V F F V   V A G V A   X A A G G   X G X D D   D A D X F
A D F F G   D G F D A   A F G A X   F F D V D   D D A G A   F A D A V
D D D A V   G A V A D   F G D D F   F D G D V   D G G X A   X A X D A
D X D V F   F X V A X   G F D A G   X F F F F   A A X D A   F V D X G
X F D A G   A G A V D   V A G A F   D G D A V   V D D D D   D F X G V
A F F A A   F F F D V   D F F A F   D A G D G   G A A A F   D X A X A
V A X D A   G A D X D   V F A F F   F G D D A   D D D F A   G D F A X
D G
```

No. 7

```
A G F G V   D D D D F   D D F X F   D D G D F   A X V D D   V D V X A
D D A X X   A A D D F   A G G F F   A X D D G   X D F A D   D F D G D
D V A X A   X F X D A   F X D D G   F X G D V   G F F G X   D A D F A
D D A F F   V D G X A   A D X F X   G V A D A   X G X A G   A G D G V
X D D V
```

No. 8

```
D F G F X   D F A F F   X D X A G   A D G G G   D D F G A   X G V D F
V V F D A   A A X G D   A V D V A   D D G V D   A F A G
```

The cryptograms have been examined for identical beginnings, and Nos. 3 and 6 apparently begin alike, identical portions being underlined as shown in figure 51. Now the number of identical sections in the two cryptograms is 15; this indicates that the width of the transposition rectangle is 15. Therefore, No. 3 (290 letters) has 5 long columns of 20 letters and 10 short columns of 19 letters [$(15\times20)-10=290$]. No. 6 (302 letters) has 2 long columns of 21 letters and 13 short columns of 20 letters [$(15\times21)-13=302$]. The identical sections in No. 3 and No. 6 having been marked off as shown in figure 51, the next step is to transcribe the texts into their correct column lengths as given by the study of identical sections, writing them merely in their serial order, as shown in figure 52. In this transcription no serious difficulty is usually encountered in the division into correct column lengths, this process being guided by the identical sequences, the number of letters between the identical sequences, and the maximum and minimum lengths of the columns as calculated from the dimensions of the rectangle. Whenever difficulties are encountered in this process, they are brought about by accidental identities of letters before and after the true or actual identical sequences. In the present case no such difficulties arise except in going from column 12 to column 13. The identical sections for column 13 here consist of the sequence A F F A A F; if these sections are placed at the head of column 13, it leaves column 12 one letter short at the bottom in each diagram. This means that the initial A's in these identical sequences represent an accidental identity; these A's belong at the bottom of column 12 in each diagram, and the true identical sequences are F F A A F, and not A F F A A F. In some cases there may be many more instances of such accidental identities before and after the true identical sequences. Another thing to be noted is that the identical beginnings in this case run along for at least 4 complete rows and part of the fifth row in the transposition rectangle. Therefore, the identical sequences should consist of not less than 4, and not more than 5 letters; any letters in excess of 5 in any identical sequence are accidental identities. There are several such accidental identities in the case under study, viz, in columns 5 and 12.

No. 3

$\underbrace{\text{X D A A}}_{1}$ A G X D D X V F F V D G A D F $\underbrace{\text{D X A A}}_{2}$ A G D F A D G

A F D A D G V G $\underbrace{\text{D V F D F}}_{3}$ X A G F X A F A F A X D D D D $\underbrace{\text{F D}}_{4}$

$\underbrace{\text{X A X}}$ V A D X F X F D G A G F G G A $\underbrace{\text{D D A G D G}}_{5}$ X A V G D G

A D A F A X F A $\underbrace{\text{A G V A}}_{6}$ A G A F D V D V D X F D A X F $\underbrace{\text{D F F}}_{7}$

$\underbrace{\text{G D}}$ X D V D A D A V D A D D D G A $\underbrace{\text{D A G A}}_{8}$ A A F G G D X A X

F G V X D D $\underbrace{\text{G D D F}}_{9}$ A F A G V A F G X G V D D A X $\underbrace{\text{X D V F F}}_{10}$

F F D X G V G D F G A V A D $\underbrace{\text{A X D A F}}_{11}$ A A F D G F V F X X X

A A G $\underbrace{\text{A G A F D G}}_{12}$ X A F A F X X G G A G A $\underbrace{\text{A F F A A F}}_{13}$ D G A

G A F V X D G G F G D $\underbrace{\text{A A A F D}}_{14}$ A D A D X V V A X F V A D D

G $\underbrace{\text{A F F F G}}_{15}$ X A X D F D D F X A A A A A

No. 6

$\underbrace{\text{X D A A}}_{1}$ V D X D G F X V G D D A V G X A $\underbrace{\text{D X A A}}_{2}$ D X G G A A

G D F D A A A G A X $\underbrace{\text{D V F D F}}_{3}$ D F F D D F D D F X F X X F D

$\underbrace{\text{F D X A X}}_{4}$ G A X F F V D V A F G V D V D $\underbrace{\text{D D A G D G}}_{5}$ G D A A

G G F D D D V F F V V $\underbrace{\text{A G V A}}_{6}$ X A A G G X G X D D D A D X F

A $\underbrace{\text{D F F G D}}_{7}$ G F D A A F G A X F F D V D D $\underbrace{\text{D A G A}}_{8}$ F A D A V

D D D A V G A V A D F $\underbrace{\text{G D D F}}_{9}$ F D G D V D G G X A X A X D A

D $\underbrace{\text{X D V F F}}_{10}$ X V A X G F D A G X F F F F A $\underbrace{\text{A X D A F}}_{11}$ V D X G

X F D A G A G A V D V $\underbrace{\text{A G A F D}}_{12}$ G D A V V D D D D D F X G V

$\underbrace{\text{A F F A A F}}_{13}$ F F D V D F F A F D A G D G G $\underbrace{\text{A A A F D}}_{14}$ X A X A

V A X D A G A D X D V F $\underbrace{\text{A F F F G}}_{15}$ D D A D D D F A G D F A X

D G

FIGURE 51.

No. 3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	D	D	F	D	A	D	D	G	X	A	A	F	A	A
D	X	V	D	D	G	F	A	D	D	X	G	F	A	F
A	A	F	X	A	V	F	G	D	V	D	A	A	A	F
A	A	D	A	G	A	G	A	F	F	A	F	A	F	F
A	A	F	X	D	A	D	A	A	F	F	D	F	D	G
G	G	X	V	G	G	X	A	F	F	A	G	D	A	X
X	D	A	A	X	A	D	F	A	F	A	X	G	D	A
D	F	G	D	A	F	V	G	G	D	F	A	A	A	X
D	A	F	X	V	D	D	G	V	X	D	F	G	D	D
X	D	X	F	G	V	A	D	A	G	G	A	A	X	F
V	G	A	X	D	D	D	X	F	V	F	F	F	V	D
F	A	F	F	G	V	A	A	G	G	V	X	V	V	D
F	F	A	D	A	D	V	X	X	D	F	X	X	A	F
V	D	F	G	D	X	D	F	G	F	X	G	D	X	X
D	A	A	A	A	F	A	G	V	G	X	G	G	F	A
G	D	X	G	F	D	D	V	D	A	X	A	G	V	A
A	G	D	F	A	A	D	X	D	V	A	G	F	A	A
D	V	D	G	X	X	D	D	A	A	A	A	G	D	A
F	G	D	G	F	F	G	D	X	D	G	A	D	D	A
		D	A	A		A							G	

FIGURE 52.

No. 6

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	D	D	F	D	A	D	D	G	X	A	A	F	A	A
D	X	V	D	D	G	F	A	D	D	X	G	F	A	F
A	A	F	X	A	V	F	G	D	V	D	A	A	A	F
A	A	D	A	G	A	G	A	F	F	A	F	A	F	F
V	D	F	X	D	X	D	F	F	F	F	D	F	D	G
D	X	D	G	G	A	G	A	D	X	V	G	F	X	D
X	G	F	A	G	A	F	D	G	V	D	D	F	A	D
D	G	F	X	D	G	D	A	D	A	X	A	D	X	A
G	A	D	F	A	G	A	V	V	X	G	V	V	A	D
F	A	D	F	A	X	A	D	D	G	X	V	D	V	D
X	G	F	V	G	G	F	D	G	F	F	D	F	A	D
V	D	D	D	G	X	G	D	G	D	D	D	F	X	F
G	F	D	V	F	D	A	A	X	A	A	D	A	D	A
D	D	F	A	D	D	X	V	A	G	G	D	F	A	G
D	A	X	F	D	D	F	G	X	X	A	D	D	G	D
A	A	F	G	D	A	F	A	A	F	G	F	A	A	F
V	A	X	V	V	D	D	V	X	F	A	X	G	D	A
G	G	X	D	F	X	V	A	D	F	V	G	D	X	X
X	A	F	V	F	F	D	D	A	F	D	V	G	D	D
A	X	D	D	V	A	D	F	D	A	V	A	G	V	G
				V									F	

FIGURE 52—Continued.

c. Now comes the attempt to place the columns in proper sequence in the respective transposition rectangles. Since No. 6 has only 2 long columns, *viz*, 5 and 14, it is obvious that these two columns belong at the extreme left of the rectangle. Their order may be 5–14 or 14–5; there is no way of telling which is correct just yet. Since No. 3 has 5 long columns, *viz*, 3, 4, 5, 7, 14, and since from No. 6 it has been ascertained that 5 and 12 go to the extreme left, it is obvious that columns 3, 4, and 7 occupy the third, fourth, and fifth positions in the rectangles. Their order may be any permutation of the three numbers 3, 4, and 7; their exact order must be ascertained by further study.

d. In this study, to fix the exact order of the columns and thus to reconstruct the transposition key, advantage can be taken of the diverse lengths of other cryptograms that may be available in the same key. In this case there are 6 additional cryptograms, Nos. 1, 2, 4, 5, 7, and 8, suitable for the purpose. The following calculations are made:

Cryptogram No.	Total number of letters	Lengths of columns	Number of columns	
			Long	Short
1	60	4	All same length	
2	38	3 and 2	8	7
4	62	5 and 4	2	13
5	74	5 and 4	14	1
7	124	9 and 8	4	11
8	54	4 and 3	9	6

Now No. 7 has 4 long columns, and these must consist of 4 columns from among the 5 already ascertained as falling at the extreme left, *viz*, 3, 4, 5, 7, and 14. Columns 5 and 14 have furthermore been placed in positions 1–2, leaving columns 3, 4, and 7 for positions 3-4-5. Which of these three possibilities is to be omitted as a long column in No. 7? A means of answering this question involves certain considerations of general importance in the cryptanalysis of this type of system.

e. Consider a transposition rectangle in which the number of columns is *even*, and consider specifically the first pair of columns in such a rectangle. The combinations of bipartite components formed by the juxtaposition of these 2 columns correspond to plain-text letters, and therefore the distribution of the bipartite digraphs in these columns will be monoalphabetic in character. The same is true with respect to the bipartite components in the third and fourth columns, the fifth and sixth columns, and so on. Hence, if a long cryptogram of this nature is at hand, and if the 2 columns which belong at the extreme left can be ascertained, then a distribution of the bipartite digraphs formed by juxtaposing these columns should not only be monoalphabetic, but also *this distribution, if it is at all normal, will afford a basis for matching other columns which will produce similar distributions*, for the text as a whole is monoalphabetic. In this way, by proper matching of columns, those which really go together to form the pairs containing the bipartite equivalents of the plain-text letters can be ascertained. From that point on, the solution of the problem is practically the same as that of solving a columnar transposition cipher with nonfractionated letters.

f. But now consider a plain-text rectangle in the ADFGVX system, in which the number of columns is *odd*, and consider specifically the first pair of columns in the rectangle. Now only the *alternate* combinations of bipartite components in these columns form the units of plain-text letters. The same is true of the bipartite components of the third and fourth, the fifth and sixth columns, and so on. In all other respects, however, the remarks contained in subparagraph *e* apply equally to this case where the width of the rectangle is odd.

g. Returning to the problem under study, it has been ascertained that columns 5 and 14 fall at the extreme left. Whether their correct order is 5–14 or 14–5 cannot at the moment be ascertained, nor is it essential. The thing to do is to make a distribution of the bipartite pairs and see what it is like. Since the width of the rectangle here is odd, only the 1st, 3d, 5th, . . . pairs down the columns can be distributed in a frequency square. The results are shown in Fig. 53.

No. 3	Col. 5	Col. 14
1	D	A
	D	A
3	A	A
	G	F
5	D	D
	G	A
7	X	D
	A	A
9	V	D
	G	X
11	D	V
	G	V
13	A	A
	D	X
15	A	F
	F	V
17	A	A
	X	D
19	F	D
	A	G

No. 6	Col. 5	Col. 14
1	D	A
	D	A
3	A	A
	G	F
5	D	D
	G	X
7	G	A
	D	X
9	A	A
	A	V
11	G	A
	G	X
13	F	D
	D	A
15	D	G
	D	A
17	V	D
	F	X
19	F	D
	V	V
21	V	F

2D COMPONENT

1ST COMPONENT	A	D	F	G	V	X
A	𝍸		/			
D	//	//		/	/	
F		///				
G	//					
V		//	/			
X		/				

FIGURE 53.

h. The distribution is fairly good. Five occurrences of **AA** are noted, 3 of **FD**. These must represent high-frequency letters. The ϕ (*Phi*) test for monoalphabeticity may be applied.

Expected value of ϕ for plain text $=.0667\times21\times20=28.01$
Expected valueof ϕ for random text $=.0385\times21\times20=16.17$
Observed value of ϕ in this case $=(5\times4)+(2\times1)+(2\times1)+(3\times2)+(2\times1)+(2\times1)=34$

The observed value of ϕ is considerably greater than the expected value for plain text and more than twice as much as the expected value for random text. Using the distribution in figure 53 as a basis, an attempt is made to add to the 5–14 combination a column selected from among columns 3, 4, and 7, so that the second, fourth, sixth . . . pairs down the second and third columns in the rectangle will give bipartite pairs that will conform to the distribution noted in figure 53. Since the results sought will be very materially affected if the combination 5–14 should really be 14–5, all possible combinations of 5–14 and 14–5 with 3, 4, and 7 must be tried. The various combinations tested are shown in figure 54.

No. 3

	(1)			(2)			(3)			(4)			(5)			(6)		
	5	14	3	5	14	4	5	14	7	14	5	3	14	5	4	14	5	7
1	D	A	D	D	A	F	D	A	D	A	D	D	A	D	F	A	D	D
2	D	A	V	D	A	D	D	A	F	A	D	V	A	D	D	A	D	F
3	A	A	F	A	A	X	A	A	F	A	A	F	A	A	X	A	A	F
4	G	F	D	G	F	A	G	F	G	F	G	D	F	G	A	F	G	G
5	D	D	F	D	D	X	D	D	D	D	D	F	D	D	X	D	D	D
6	G	A	X	G	A	V	G	A	X	A	G	X	A	G	V	A	G	X
7	X	D	A	X	D	A	X	D	D	D	X	A	D	X	A	D	X	D
8	A	A	G	A	A	D	A	A	V	A	A	G	A	A	D	A	A	V
9	V	D	F	V	D	X	V	D	D	D	V	F	D	V	X	D	V	D
10	G	X	X	G	X	F	G	X	A	X	G	X	X	G	F	X	G	A
11	D	V	A	D	V	X	D	V	D	V	D	A	V	D	X	V	D	D
12	G	V	F	G	V	F	G	V	A	V	G	F	V	G	F	V	G	A
13	A	A	A	A	A	D	A	A	V	A	A	A	A	A	D	A	A	V
14	D	X	F	D	X	G	D	X	D	X	D	F	X	D	G	X	D	D
15	A	F	A	A	F	A	A	F	A	F	A	A	F	A	A	F	A	A
16	F	V	X	F	V	G	F	V	D	V	F	X	V	F	G	V	F	D
17	A	A	D	A	A	F	A	A	D	A	A	D	A	A	F	A	A	D
18	X	D	D	X	D	G	X	D	D	D	X	D	D	X	G	D	X	D
19	F	D	D	F	D	G	F	D	G	D	F	D	D	F	G	D	F	G
20	A	G	D	A	G	A	A	G	A	G	A	D	G	A	A	G	A	A

No. 6

	(1)			(2)			(3)			(4)			(5)			(6)		
	5	14	3	5	14	4	5	14	7	14	5	3	14	5	4	14	5	7
1	D	A	D	D	A	F	D	A	D	A	D	D	A	D	F	A	D	D
2	D	A	V	D	A	D	D	A	F	A	D	V	A	D	D	A	D	F
3	A	A	F	A	A	X	A	A	F	A	A	F	A	A	X	A	A	F
4	G	F	D	G	F	A	G	F	G	F	G	D	F	G	A	F	G	G
5	D	D	F	D	D	X	D	D	D	D	D	F	D	D	X	D	D	D
6	G	X	D	G	X	G	G	X	G	X	G	D	X	G	G	X	G	G
7	G	A	F	G	A	A	G	A	F	A	G	F	A	G	A	A	G	F
8	D	X	F	D	X	X	D	X	D	X	D	F	X	D	X	X	D	D
9	A	A	D	A	A	F	A	A	A	A	A	D	A	A	F	A	A	A
10	A	V	D	A	V	F	A	V	A	V	A	D	V	A	F	V	A	A
11	G	A	F	G	A	V	G	A	F	A	G	F	A	G	V	A	G	F
12	G	X	D	G	X	D	G	X	G	X	G	D	X	G	D	X	G	G
13	F	D	D	F	D	V	F	D	A	D	F	D	D	F	V	D	F	A
14	D	A	F	D	A	A	D	A	X	A	D	F	A	D	A	A	D	X
15	D	G	X	D	G	F	D	G	F	G	D	X	G	D	F	G	D	F
16	D	A	F	D	A	G	D	A	F	A	D	F	A	D	G	A	D	F
17	V	D	X	V	D	V	V	D	D	D	V	X	D	V	V	D	V	D
18	F	X	X	F	X	D	F	X	V	X	F	X	X	F	D	X	F	V
19	F	D	F	F	D	V	F	D	D	D	F	F	D	F	V	D	F	D
20	V	V	D	V	V	D	V	V	D	V	V	D	V	V	D	V	V	D
21	V	F		V	F		V	F		F	V		F	V		F	V	

FIGURE 54.

i. Frequency distributions are now made. If combination 5–14–3 is correct for No. 3, it is also correct for No. 6. Hence, a single distribution is made of the bipartite pairs in rows 1, 3, 5, . . . of columns 5–14, and of the pairs in rows 2, 4, 6, . . . of columns 14–3. Similar distributions are made of the pairs given under each of the other combinations. These distributions are shown in figure 55.

(1) [5–14–3]

	A	D	F	G	V	X
A	𝍸		///	/	//	/
D	//	///		/	/	
F		𝍸				
G	//	/				
V		////	//			/
X		///	//			//

(2) [5–14–4]

	A	D	F	G	V	X
A	𝍸 /	///	/	/	/	
D	//	//		//	/	
F	//	///				
G	///					
V		///	///	/		
X		///	/	//		/

(3) [5–14–7]

	A	D	F	G	V	X
A	𝍸		////		/	//
D	//	///		/	/	
F		///		//		
G	///					
V	//	////	/			
X	/	///		//	/	

(4) [14–5–3]

	A	D	F	G	V	X
A	𝍸	////		///		
D		//	𝍸 //		////	/
F	/				/	//
G		𝍸	/			//
V		//				
X		/				

(5) [14–5–4]

	A	D	F	G	V	X
A	𝍸 /	///	/	//		
D	/	////	///	//	//	//
F	/	/		/	/	
G	//	//	//	/	/	
V		//				
X				/		

(6) [14–5–7]

	A	D	F	G	V	X
A	𝍸 //	//		//	/	
D		////	𝍸 /		//	//
F	/	/			//	
G	//	/		////		/
V		//				
X		/				

Figure 55.

j. These distributions are now tested for monoalphabeticity, by applying the ϕ test. The number of occurrences in each distribution is 41. Then $41 \times 40 \times .0667 = 109.4$ is the expected value of ϕ for plain text; $41 \times 40 \times .0385 = 63.1$ is the expected value of ϕ for random text. Here are the calculations for the first distribution (combination 5–14–3) yielding the observed value of ϕ as 82:

$$(5\times4)+(3\times2)+(1\times0)+(2\times1)+(1\times0)+$$
$$(2\times1)+(3\times2)+(1\times0)+(1\times0)+(5\times4)+(2\times1)+$$
$$(1\times0)+(4\times3)+(2\times1)+(1\times0)+(3\times2)+(2\times2)+(2\times1)=82.$$

The observed values of ϕ for all 6 frequency distributions are shown herewith:

(1)=82 (4)=120
(2)=76 (5)= 70
(3)=78 (6)=110

Only two of these distributions give close approximations to 109, the expected value of ϕ, and they may be retained for further experiment. They are the ones for combinations (4) and (6), with values of 120 and 110, respectively.

k. Selecting combinations (4) and (6) *viz*, 14–5–3, and 14–5–7, since columns 14, 3, 4, 5 and 7 form the group of 5 columns at the left of the transposition rectangle, the following combinations are possible:

(1)	14–5–3–4–7	(3)	14–5–7–3–4
(2)	14–5–3–7–4	(4)	14–5–7–4–3

l. The following sets of columns correspond to these 4 combinations in the 2 cryptograms (fig. 56):

No. 3

	(1)					(2)					(3)					(4)				
	14	5	3	4	7	14	5	3	7	4	14	5	7	3	4	14	5	7	4	3
1	A	D	D	F	D	A	D	D	D	F	A	D	D	D	F	A	D	D	F	D
2	A	D	V	D	F	A	D	V	F	D	A	D	F	V	D	A	D	F	D	V
3	A	A	F	X	F	A	A	F	F	X	A	A	F	F	X	A	A	F	X	F
4	F	G	D	A	G	F	G	D	G	A	F	G	G	D	A	F	G	G	A	D
5	D	D	F	X	D	D	D	F	D	X	D	D	D	F	X	D	D	D	X	F
6	A	G	X	V	X	A	G	X	X	V	A	G	X	X	V	A	G	X	V	X
7	D	X	A	A	D	D	X	A	D	A	D	X	D	A	A	D	X	D	A	A
8	A	A	G	D	V	A	A	G	V	D	A	A	V	G	D	A	A	V	D	G
9	D	V	F	X	D	D	V	F	D	X	D	V	D	F	X	D	V	D	X	F
10	X	G	X	F	A	X	G	X	A	F	X	G	A	X	F	X	G	A	F	X
11	V	D	A	X	D	V	D	A	D	X	V	D	D	A	X	V	D	D	X	A
12	V	G	F	F	A	V	G	F	A	F	V	G	A	F	F	V	G	A	F	F
13	A	A	A	D	V	A	A	A	V	D	A	A	V	A	D	A	A	V	D	A
14	X	D	F	G	D	X	D	F	D	G	X	D	D	F	G	X	D	D	G	F
15	F	A	A	A	A	F	A	A	A	A	F	A	A	A	A	F	A	A	A	A
16	V	F	X	G	D	V	F	X	D	G	V	F	D	X	G	V	F	D	G	X
17	A	A	D	F	D	A	A	D	D	F	A	A	D	D	F	A	A	D	F	D
18	D	X	D	G	D	D	X	D	D	G	D	X	D	D	G	D	X	D	G	D
19	D	F	D	G	G	D	F	D	G	G	D	F	G	D	G	D	F	G	G	D
20	G	A	D	A	A	G	A	D	A	A	G	A	A	D	A	G	A	A	A	D

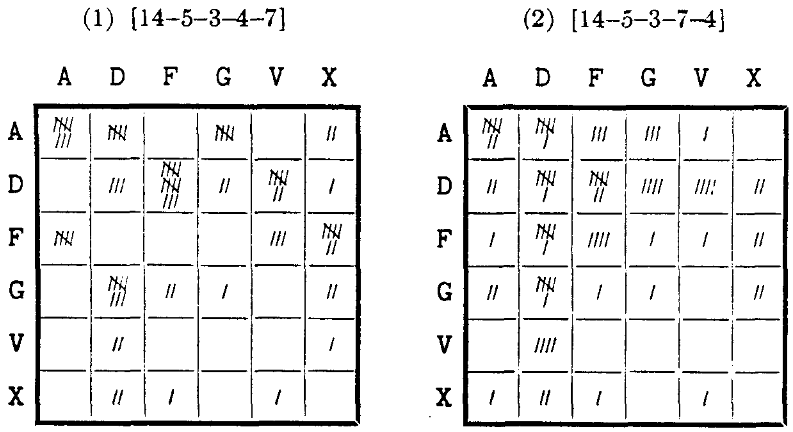
FIGURE 56.

No. 6

	(1)					(2)					(3)					(4)				
	14	5	3	4	7	14	5	3	7	4	14	5	7	3	4	14	5	7	4	3
1	A	D	D	F	D	A	D	D	D	F	A	D	D	D	F	A	D	D	F	D
2	A	D	V	D	F	A	D	V	F	D	A	D	F	V	D	A	D	F	D	V
3	A	A	F	X	F	A	A	F	F	X	A	A	F	F	X	A	A	F	X	F
4	F	G	D	A	G	F	G	D	G	A	F	G	G	D	A	F	G	G	A	D
5	D	D	F	X	D	D	D	F	D	X	D	D	D	F	X	D	D	D	X	F
6	X	G	D	G	G	X	G	D	G	G	X	G	G	D	G	X	G	G	G	D
7	A	G	F	A	F	A	G	F	F	A	A	G	F	F	A	A	G	F	A	F
8	X	D	F	X	D	X	D	F	D	X	X	D	D	F	X	X	D	D	X	F
9	A	A	D	F	A	A	A	D	A	F	A	A	A	D	F	A	A	A	F	D
10	V	A	D	F	A	V	A	D	A	F	V	A	A	D	F	V	A	A	F	D
11	A	G	F	V	F	A	G	F	F	V	A	G	F	F	V	A	G	F	V	F
12	X	G	D	D	G	X	G	D	G	D	X	G	G	D	D	X	G	G	D	D
13	D	F	D	V	A	D	F	D	A	V	D	F	A	D	V	D	F	A	V	D
14	A	D	F	A	X	A	D	F	X	A	A	D	X	F	A	A	D	X	A	F
15	G	D	X	F	F	G	D	X	F	F	G	D	F	X	F	G	D	F	F	X
16	A	D	F	G	F	A	D	F	F	G	A	D	F	F	G	A	D	F	G	F
17	D	V	X	V	D	D	V	X	D	V	D	V	D	X	V	D	V	D	V	X
18	X	F	X	D	V	X	F	X	V	D	X	F	V	X	D	X	F	V	D	X
19	D	F	F	V	D	D	F	F	D	V	D	F	D	F	V	D	F	D	V	F
20	V	V	D	D	D	V	V	D	D	D	V	V	D	D	D	V	V	D	D	D
21	F	V				F	V				F	V				F	V			

FIGURE 56—Continued.

m. The additional bipartite pairs given by adding columns 4–7 to the basic combination 14–5–3 are distributed in the 4th frequency distribution square of figure 55, yielding the distribution shown in square (1) of figure 57. The other squares in figure 57 are constructed in the same way, for the other combinations of figure 56.


(1) [14–5–3–4–7]

	A	D	F	G	V	X
A	𝍸///	𝍸		𝍸		//
D		///	𝍸𝍸///	//	𝍸//	/
F	𝍸				///	𝍸//
G		𝍸///	//	/		//
V		//				/
X		//	/		/	

(2) [14–5–3–7–4]

	A	D	F	G	V	X
A	𝍸//	𝍸/	///	///	/	
D	//	𝍸/	𝍸//	////	////	//
F	/	𝍸/	////	/	/	//
G	//	𝍸/	/	/		//
V		////				
X	/	//	/		/	

FIGURE 57.

(3) [14–5–7–3–4]

	A	D	F	G	V	X
A	𝍸///	////		//	/	
D	𝍸	𝍸////	𝍸𝍸/	//	//	///
F	//	/	𝍸	//	//	//
G	//	///		////		/
V	/	////				
X		//	/	/	/	

(4) [14–5–7–4–3]

	A	D	F	G	V	X
A	𝍸///	𝍸	//	//	//	
D	/	𝍸/	𝍸////	/	𝍸/	𝍸//
F	//	//	//		///	///
G	//	///	//	𝍸		//
V		///				/
X		/	/			

FIGURE 57—Continued.

n. Again applying the ϕ-test, the expected value of ϕ is $81 \times 80 \times .0667 = 432$. The observed values for the four combinations of figure 57 are as follows:

(1) For combination 14–5–3–4–7, $\phi = 436$
(2) For combination 14–5–3–7–4, $\phi = 276$
(3) For combination 14–5–7–3–4, $\phi = 344$
(4) For combination 14–5–7–4–3, $\phi = 318$

The combination 14–5–3–4–7, giving the greatest value for ϕ (a little better than the expected value), is very probably the correct one.

o. Examining the other cryptograms that are available, it is seen that No. 7 is the third longest one of the entire set, with 124 letters; moreover, the dimensions of the rectangle $[(15 \times 9) - 11 = 124]$ are such as to bring about 4 long columns of 9 letters and 11 columns of 8 letters. The first 5 columns are definitely fixed in position, since it is known that the first 5 key numbers are 14–5–3–4–7. The resulting diagram is shown in figure 58. There is now a section consisting of

14	5	3	4	7	1	2	6	8	9	10	11	12	13	15
A	X	D	V	D	A	D	F	D	X	F	G	D	A	G
D	A	G	D	F	G	F	F	D	D	X	X	A	A	D
A	A	D	V	A	F	D	A	V	A	G	D	F	D	G
X	D	F	X	D	G	D	X	A	F	D	A	F	X	V
G	D	A	A	D	V	F	D	X	X	V	D	V	F	X
X	F	X	D	F	D	X	D	A	D	G	F	D	X	D
A	A	V	D	D	D	F	G	X	D	F	A	G	G	D
G	G	D	A	G	D	D	X	F	G	F	D	X	V	V
A	G	D	X											

FIGURE 58.

10 columns which are to be anagrammed to ascertain their correct sequence. The column to follow column 7 is ascertained on the basis of the repetitions which are brought about when the selected column is placed on the right. These repetitions should fall into those cells of frequency distribution (1), figure 57, which are of high frequency. In other words, the process is one of selecting from among columns 1, 2, 6, 8, 9, 10, 11, 12, 13, and 15 that column which will yield the most repetitions of bipartite digraphs with the digraphs given by the juxtaposition of columns 14–5–3–4–7, as distributed in frequency square (1) of figure 57. The column thus selected turns out to be No. 10. Then other columns are added by proceeding along the same lines, the work becoming progressively more easy as the number of available candidates decreases. Sometimes the discovery of what appears to be a long repetition within one of the cryptograms or between two cryptograms facilitates the process. In this case the results obtained from the 3 cryptograms under study are shown in figure 59.

No. 3

14	5	3	4	7	10	15	12	13	1	2	8	6	9	11
A	D	D	F	D	X	A	A	F	X	D	D	A	G	A
A	D	V	D	F	D	F	G	F	D	X	A	G	D	X
A	A	F	X	F	V	F	A	A	A	A	G	V	D	D
F	G	D	A	G	F	F	F	A	A	A	A	A	F	A
D	D	F	X	D	F	G	D	F	A	A	A	A	A	F
A	G	X	V	X	F	X	G	D	G	G	A	G	F	A
D	X	A	A	D	F	A	X	G	X	D	F	A	A	A
A	A	G	D	V	D	X	A	A	D	F	G	F	G	F
D	V	F	X	D	X	D	F	G	D	A	G	D	V	D
X	G	X	F	A	G	F	A	A	X	D	D	V	A	G
V	D	A	X	D	V	D	F	F	V	G	X	D	F	F
V	G	F	F	A	G	D	X	V	F	A	A	V	G	V
A	A	A	D	V	D	F	X	X	F	F	X	D	X	F
X	D	F	G	D	F	X	G	D	V	D	F	X	G	X
F	A	A	A	A	G	A	G	G	D	A	G	F	V	X
V	F	X	G	D	A	A	A	G	G	D	V	D	D	X
A	A	D	F	D	V	A	G	F	A	G	X	A	D	A
D	X	D	G	D	A	A	A	G	D	V	D	X	A	A
D	F	D	G	G	D	A	A	D	F	G	D	F	X	G
G	A	D	A	A										

FIGURE 59.

No. 6

14	5	3	4	7	10	15	12	13	1	2	8	6	9	11
A	D	D	F	D	X	A	A	F	X	D	D	A	G	A
A	D	V	D	F	D	F	G	F	D	X	A	G	D	X
A	A	F	X	F	V	F	A	A	A	A	G	V	D	D
F	G	D	A	G	F	F	F	A	A	A	A	A	F	A
D	D	F	X	D	F	G	D	F	V	D	F	X	F	F
X	G	D	G	G	X	D	G	F	D	X	A	A	D	V
A	G	F	A	F	V	D	D	F	X	G	D	A	G	D
X	D	F	X	D	A	A	A	D	D	G	A	G	D	X
A	A	D	F	A	X	D	V	V	G	A	V	G	V	G
V	A	D	F	A	G	D	V	D	F	A	D	X	D	X
A	G	F	V	F	F	D	D	F	X	G	D	G	G	F
X	G	D	D	G	D	F	D	F	V	D	D	X	G	D
D	F	D	V	A	A	A	D	A	G	F	A	D	X	A
A	D	F	A	X	G	G	D	F	D	D	V	D	A	G
G	D	X	F	F	X	D	D	D	D	A	G	D	X	A
A	D	F	G	F	F	F	F	A	A	A	A	A	A	G
D	V	X	V	D	F	A	X	G	V	A	V	D	X	A
X	F	X	D	V	F	X	G	D	G	G	A	X	D	V
D	F	F	V	D	F	D	V	G	X	A	D	F	A	D
V	V	D	D	D	A	G	A	G	A	X	F	A	D	V
F	V													

No. 7

14	5	3	4	7	10	15	12	13	1	2	8	6	9	11
A	X	D	V	D	F	G	D	A	A	D	D	F	X	G
D	A	G	D	F	X	D	A	A	G	F	D	F	D	X
A	A	D	V	A	G	G	F	D	F	D	V	A	A	D
X	D	F	X	D	D	V	F	X	G	D	A	X	F	A
G	D	A	A	D	V	X	V	F	V	F	X	D	X	D
X	F	X	D	F	G	D	D	X	D	X	A	D	D	F
A	A	V	D	D	F	D	G	G	D	F	X	G	D	A
G	G	D	A	G	F	V	X	V	D	D	F	X	G	D
A	G	D	X											

Figure 59—Continued.

p. What the cryptanalyst now has before him is a monoalphabetic substitution cipher, the solution of which presents no difficulties. The cipher square is reconstructed as completely as possible, blanks being left where there are no occurrences to give clues as to the character involved, usually some of the digits and the very infrequent letters. In this case the only letters which do not occur in the plain text are Q, X, and Z. The digits 5 and 7 are recovered from the context, in message No. 6, where the caliber of a gun is mentioned and the digits are confirmed at other places in the message. The square that is obtained is seen in figure 60. Examination of the mixed sequence discloses that it is based upon the phrase THE FLOWERS THAT BLOOM IN THE SPRING. This permits of the establishment of the transposition key and of the position of the digits in the checkerboard (as in par. 38*i*). The results are shown in figure 61. The completely solved messages are shown in figure 62.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15											
Literal key	T	H	E	F	L	O	W	R	S	A	B	M	I	N	P	G	C	D	J	K	Q	U	V	X	Y	Z
Numerical key	14	5	3	4	7	10	15	12	13	1	2	8	6	9	11											

2D COMPONENT

1ST COMPONENT

	A	D	F	G	V	X
A	T	H		E	5	F
D		L	O	W	R	S
F	A		B		M	I
G		N	P	G	7	C
V		D		J		K
X		U	V		Y	

FIGURE 60.

2D COMPONENT

1ST COMPONENT

	A	D	F	G	V	X
A	T	H	8	E	5	F
D	6	L	O	W	R	S
F	A	1	B	2	M	I
G	9	N	P	G	7	C
V	3	D	4	J	0	K
X	Q	U	V	X	Y	Z

FIGURE 61.

No. 1

14	5	3	4	7	10	15	12	13	1	2	8	6	9	11
R		E		G		I		M		E		N		T
D	V	A	G	G	G	F	X	F	V	A	G	G	D	A
	I		N		P		O		S		I		T	
A	F	X	G	D	G	F	D	F	D	X	F	X	A	A
I		O		N		S		H		A		L		L
F	X	D	F	G	D	D	X	A	D	F	A	D	D	D
	I		A		T		T		A		C		K	
D	F	X	F	A	A	A	A	A	F	A	G	X	V	X

No. 2

14	5	3	4	7	10	15	12	13	1	2	8	6	9	11
R		E		Q		U		E		S		T		I
D	V	A	G	X	A	X	D	A	G	D	X	A	A	F
	N		S		T		R		U		C		T	
X	G	D	D	X	A	A	D	V	X	D	G	X	A	A
I		O		N		S								
F	X	D	F	G	D	D	X							

FIGURE 62.

No. 3

```
14 5   3  4   7  10  15 12  13 1   2  8   6  9   11
---------------------------------------------------
  H      O      S      T      I      L      E     T
A  D   D  F   D  X   A  A   F  X   D  D   A  G   A

     R      O      O      P      S      E      S
A  D   V  D   F  D   F  G   F  D   X  A   G  D   X

  T      I      M      A      T      E      D     O
A  A   F  X   F  V   F  A   A  A   A  G   V  D   D

     N      E      B      A      T      T      A
F  G   D  A   G  F   F  F   A  A   A  A   A  F   A

  L      I      O      N      A      T      T     A
D  D   F  X   D  F   G  D   F  A   A  A   A  A   F

     C      K      I      N      G      E      A
A  G   X  V   X  F   X  G   D  G   G  A   G  F   A

  S      T      O      F      C      O      T     T
D  X   A  A   D  F   A  X   G  X   D  F   A  A   A

     E      R      S      T      O      P      P
A  A   G  D   V  D   X  A   A  D   F  G   F  G   F

  R      I      S      O      N      E      R     S
D  V   F  X   D  X   D  F   G  D   A  G   D  V   D

     C      A      P      T      U      R      E
X  G   X  F   A  G   F  A   A  X   D  D   V  A   G

  D      F      R      O      M      C      O     M
V  D   A  X   D  V   D  F   F  V   G  X   D  F   F

     P      A      N      Y      A      5      7
F  G   F  F   A  G   D  X   V  F   A  A   V  G   V

  T      H      D      I      V      I      S     I
A  A   A  D   V  D   F  X   X  F   F  X   D  X   F

     O      N      I      N      D      I      C
X  D   F  G   D  F   X  G   D  V   D  F   X  G   X

  A      T      E      E      N      E      M     Y
F  A   A  A   A  G   A  G   G  D   A  G   F  V   X

     I      N      T      E      N      D      S
V  F   X  G   D  A   A  A   G  G   D  V   D  D   X

  T      O      R      E      A      C      H     H
A  A   D  F   D  V   A  G   F  A   G  X   A  D   A

     U      N      T      E      R      S      T
D  X   D  G   D  A   A  A   G  D   V  D   X  A   A

  O      W      N      T      O      N      I     G
D  F   D  G   G  D   A  A   D  F   G  D   F  X   G

     H      T
G  A   D  A   A
```

FIGURE 62—Continued.

No. 4

```
14 5    3 4     7 10    15 12   13 1    2 8     6 9     11
 T       H       I       R       T       Y       S       I
A A     A D     F X     D V     A A     X V     D X     F

     X       T       H       F       A       L       E
X X     G A     A A     D A     X F     A D     D A     G

 A       V       I       N       G       G       O       L
F A     X F     F X     G D     G G     G G     D F     D

     D       E       N       V       I       L       L
D V     D A     G G     D X     F F     X D     D D     D

 E
A G
```

No. 5

```
14 5    3 4     7 10    15 12   13 1    2 8     6 9     11
 C       O       R       P       S       W       I       L
G X     D F     D V     G F     D X     D G     F X     D

     L       T       A       K       E       O       V
D D     D A     A F     A V     X A     G D     F X     F

 E       R       T       R       A       F       F       I
A G     D V     A A     D V     F A     A X     A X     F

     C       C       O       N       T       R       O
X G     X G     X D     F G     D A     A D     V D     F

 L       A       T       O       N       C       E
D D     F A     A A     D F     G D     G X     A G
```

Figure 62—Continued.

No. 6

```
14 5  3 4   7 10  15 12 13 1   2 8   6 9   11
 H     O     S     T     I     L     E     T
A D   D F   D X   A A   F X   D D   A G   A
    R     O     O     P     S     E     S
A D   V D   F D   F G   F D   X A   G D   X
 T     I     M     A     T     E     D     O
A A   F X   F V   F A   A A   A G   V D   D
    N     E     B     A     T     T     A
F G   D A   G F   F F   A A   A A   A F   A
 L     I     O     N     M     O     V     I
D D   F X   D F   G D   F V   D F   X F   F
    N     G     U     P     S     T     R
X G   D G   G X   D G   F D   X A   A D   V
 E     A     M     L     I     N     E     S
A G   F A   F V   D D   F X   G D   A G   D
    O     U     T     H     W     E     S
X D   F X   D A   A A   D D   G A   G D   X
 T     O     F     R     J     5     7     7
A A   D F   A X   D V   V G   A V   G V   G
    H     A     N     D     A     S     S
V A   D F   A G   D V   D F   A D   X D   X
 E     M     B     L     I     N     G     I
A G   F V   F F   D D   F X   G D   G G   F
    N     W     O     O     D     S     N
X G   D D   G D   F D   F V   D D   X G   D
 O     R     T     H     E     A     S     T
D F   D V   A A   A D   A G   F A   D X   A
    O     F     G     O     L     D     E
A D   F A   X G   G D   F D   D V   D A   G
 N     V     I     L     L     E     S     T
G D   X F   F X   D D   D D   A G   D X   A
    O     P     B     A     T     T     E
A D   F G   F F   F F   A A   A A   A A   G
 R     Y     O     F     7     5     S     F
D V   X V   D F   A X   G V   A V   D X   A
    I     R     I     N     G     F     R
X F   X D   V F   X G   D G   G A   X D   V
 O     M     O     R     C     H     A     R
D F   F V   D F   D V   G X   A D   F A   D
    D     L     E     E     F     A     R
V V   D D   D A   G A   G A   X F   A D   V
 M
F V
```

FIGURE 62—Continued.

No. 7

```
14 5  3 4   7 10  15 12 13 1  2 8   6 9   11
_____________________________________________
 F     R     O     N     T     L     I     N
A X   D V   D F   G D   A A   D D   F X   G

    E     O     U     T     P     O     S
D A   G D   F X   D A   A G   F D   F D   X

 T     R     E     P     O     R     T     S
A A   D V   A G   G F   D F   D V   A A   D

    O     U     R     I     N     F     A
X D   F X   D D   V F   X G   D A   X F   A

 N     T     R     Y     M     I     S     S
G D   A A   D V   X V   F V   F X   D X   D

    I     O     N     S     S     H     O
X F   X D   F G   D D   X D   X A   D D   F

 T     D     O     W     N     I     N     E
A A   V D   D F   D G   G D   F X   G D   A

    N     E     M     Y     L     I     N
G G   D A   G F   V X   V D   D F   X G   D

 E     S
A G   D X
```

No. 8

```
4 5   3 4   7 10  15 12 13 1  2 8   6 9   11
_____________________________________________
 W     I     R     E     L     I     N     E
D G   F X   D V   A G   D D   F X   G D   A

    T     O     B     R     I     G     A
G A   A D   F F   F D   V F   X G   G F   A

 D     I     N     T     E     R     R     U
V D   F X   G D   A A   A G   D V   D V   X

    P     T     E     D
D G   F A   A A   G V   D
```

FIGURE 62—Continued.

40. Special solution by the exact factor method.—*a*. The student who has comprehended the successive steps in the solution of the example discussed in the preceding paragraph is in a position to grasp at once the mechanics of the special solution by the exact factor method. The latter is based upon the interception of a number of cryptograms, preferably lengthy ones, which have been enciphered by rectangles in which the last row is completely filled with letters. The total number of bipartite components in the case of such a cryptogram will yield clues as to the dimensions of the transposition rectangle. Then the text is transcribed into columns of appropriate length, all being equal in this respect, and the process of combining columns, as explained in paragraph 39*e*, is applied in order to produce the best monoalphabetic distribution of bipartite digraphs down the juxtaposed columns. There is nothing to prevent the simultaneous use of all cryptograms that have been enciphered by completely filled rectangles, for it is clear that if, for example, columns 15 and 4 are to be paired in one cryptogram, the same columns will be paired in all the other cryptograms. Hence, even if the rectangles are small in depth they can be used in this process; it is necessary only that all columns of any rectangle be of the same length. Now if only two or three such pairs of columns can be set up correctly, solution follows almost as a matter of course. No additional or new principles need be brought into play, beyond those already possessed by the student.

b. In this special solution, the important step is, of course, the initial one of experimenting with rectangles of various dimensions until the correct size has been hit upon. In some cases, excessive experimentation may not be necessary if the total number of characters is such as to yield only one or two possibilities with regard to the length of the columns. For example, suppose that previous work has established the fact that the enemy uses transposition rectangles not less than 15 and not more than 22 columns in width. A message totaling 703 letters would indicate a rectangle of 19 columns of 37 letters, since these two numbers are the only factors of 703. If this then were corroborated by other cryptograms of 76 (19×4), 152 (19×8), 190 (19×10) letters, the probability that 19 is the width of the transposition rectangle becomes quite persuasive. Of course, there will be and there should be other cryptograms of lengths that do not factor exactly; these represent the ones in which the rectangles are not completely filled in their last row. They do not enter into the solution at first, but just as soon as the positions of two or three key numbers become fixed, the data afforded by these messages become available for use in the later stages in the solution.

c. The exact-factor method is a useful one to know. For despite all instructions that may be drawn up insisting upon the advisability of not completing the last row of a transposition rectangle, the tendency to violate such a rule is quite marked, especially where a large cryptographic personnel must be employed. It is not astonishing to find that for lazy or ignorant clerks the temptation to fill the rectangle completely is particularly hard to resist when it happens that a message falls just one, two, or three letters short of forming a completely-filled rectangle: it is so much easier for such clerks to handle a rectangle with equal-length columns than one in which this is not the case. Moreover, the number of errors and therefore the number of times a shiftless or careless clerk must go over his work to correct errors is reduced to a minimum. Hence, it often happens that in such cases an enciphering clerk adds one, two, or three letters to complete the last row, thus leading to the transmission of not a few cryptograms enciphered by completely-filled rectangles. Space forbids giving an example of such a solution.

41. General solution for the ADFGVX system.—*a.* All three of the foregoing methods of solving cryptograms in the ADFGVX system fall in the category of special solutions and therefore are dependent upon the fortuitous existence of the special conditions required under each case. What is really desired in the practical situation is a method of solution which is not so dependent upon chance or good fortune for success. A search for a general solution was, of course, made during the time that the system was under minute study by the cryptanalytic agencies of the Allies, but no general solution was devised. All the solutions made during actual hostilities and for a number of weeks thereafter were of the special types described in the preceding paragraphs. The first published description of a general solution is to be found in Givierge's *Cours de Cryptographie*, 1925, but only in broad outlines. A complete general solution was independently conceived by a group of cryptanalysts in the office of the Chief Signal Officer [5] and will be described in paragraphs 42 and 43.

b. The attention of the student is directed to the comments made in paragraph 18, with regard to the significance of the term *general solution* in cryptanalysis. He must be cautioned not to expect that in practical work a general solution will, in the cryptanalytic as in the mathematical field, *invariably* lead to a solution. If there is a sufficient amount of text and if the text contains no abnormalities, the attempt to apply the general solution will usually be successful. But the cryptanalyst must remember that the ADFGVX system is by no means a simple one to solve even under the best of conditions and if there is only a small amount of text, if it happens that the transposition key is unusually long, or if the text is abnormal, he may not succeed in solving the messages by the straightforward method to be set forth below, and he may have to introduce special modifications. For the latter he can only rely upon his own ingenuity and intuition.

---

[5] See footnote 7 of this section.

42. Basic principles of the general solution.—*a*. Every transposition rectangle in the ADFGVX system must conform to one or the other of two and only two fundamental types: the number of columns must be either odd or even. A number of important consequences follow from this simple fact, some of which have already been pointed out in paragraph 39*e*. They will be elaborated upon in the next subparagraphs.

b. Consider a rectangle with an even number of columns. Each of its rows contains an even number of bipartite components, half of which are *initial* components, half, *final* components, alternating in a regular order from left to right in the rows. When the transposition is applied, all the components within a given column are of the same class, either initial or final. No intermixture or alternation of the two classes is possible. On the other hand, consider a rectangle with an odd number of columns. Each of its rows contains an odd number of bipartite components, the 1st row containing one more initial component than final components, the 2d row containing one more final component than initial components, and so on, this arrangement alternating regularly in the successive rows of the rectangle. When one studies the various columns of the rectangle, it is seen that in each column there is a perfectly regular alternation of initial and final components, the odd columns (1st, 3d, 5th, . . .) beginning with an initial component, the even columns (2d, 4th, 6th, . . .) beginning with a final component. This alternation in components remains true even after the transposition is applied. These remarks become very clear if one studies figure 63. Two transposition rectangles are shown, one with an even number of columns, the other with an odd number. Instead of the actual components (ADFGVX), the symbols θ_1 and θ_2 are used to indicate the two classes of components, initial and final, because in this analysis interest centers not upon the actual identity of a component but upon the *class* to which it belongs, initial or final. At the top of each column is placed a "plus" to denote a column occupying an odd-numbered position in the rectangle, or a "minus" to denote a column occupying an even-numbered position.

EVEN NUMBER OF COLUMNS

+	−	+	−	+	−	+	−	+	−
θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2

a

ODD NUMBER OF COLUMNS

+	−	+	−	+	−	+	−	+
θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1
θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1

b

FIGURE 63.

c. In what follows, the term "odd column" will mean merely that the column in question occupies an odd position (1st, 3d, 5th, . . .) in the transposition rectangle; the term "even column," that it occupies an even position (2d, 4th, 6th, . . .) in the rectangle. The odd or even designation has no reference whatever to the nature of the transposition key number applicable to that column, whether it is odd or even. Now when the transposition is applied to the even-width rectangle *a*, figure 63, the cryptographic text will consist of a number of sections of letters, each section corresponding to a column of the rectangle, and therefore the number of sections in this case will be even. Moreover, all the components in a section corresponding to an odd column in rectangle *a* will be Θ_1 or initial components, all those in a section corresponding to an even column, Θ_2 or final components. The sections or columns are completely homogeneous with respect to the class to which their constituent components belong. On the other hand, when the transposition is applied to odd-width rectangle *b*, the cryptographic text will consist of an odd number of sections, each corresponding to a column of the rectangle. The components in the sections consist of members of both classes of components in a regular alternation; in a section corresponding to an odd column the order is $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$ . . .; in a section corresponding to an even column the order is $\Theta_2 \rightarrow \Theta_1 \rightarrow \Theta_2$ . . .. The sections or columns are not homogeneous in this case as they are in the former.

d. Now if there were some way of distinguishing between initial components as a class and final components as a class it is clear that it may be possible first of all to ascertain whether the transposition rectangle contains an even or an odd number of columns. Secondly it may be possible to identify those columns which are even and those which are odd. Finally, it may be possible to ascertain which are the long columns and which are short, thus yielding the exact outlines of the rectangle in case the last row is incompletely filled. From that point on, solution follows along the same lines as explained in paragraph 40, with the modification that in the pairing of columns the number of possibilities is greatly reduced, since it is useless to pair two columns both containing initial components or final components.

e. The foregoing depends then upon the possibility of being able to distinguish as a class between initial and final components of the bipartite cipher equivalents in this system, or at least between letters belonging to one or the other of these two general classes of components. Now if the substitution checkerboard has not been consciously manipulated with a view to destroying certain properties normally characterizing its rows and columns, the sort of differentiation indicated above is quite possible. For example, if in the checkerboard shown in figure 61 the normal frequencies of the letters as they appear in English telegraphic plain text [6] are inserted in the cells and totals are obtained vertically and horizontally, these totals will permit of assigning frequency weights to the letters ADFGVX as initial and as final letters of the bipartite cipher equivalents of the plain-text letters. This is shown below in figure 64. The bipartite letter A

	2D COMPONENT						
1ST COMPONENT	A	D	F	G	V	X	Sums
A	T 92	H 34		E 130		F 28	284
D		L 36	O 75	W 16	R 76	S 61	264
F	A 74		B 10		M 25	I 74	183
G		N 79	P 27	G 16		C 31	153
V		D 42		J 2		K 3	47
X	Q 3	U 26	V 15	X 5	Y 19	Z 1	69
Sums	169	217	127	169	120	198	1,000

FIGURE 64.

---

[6] As given in fig. 3, p. 13, *Military Cryptanalysis, Part 1.*

has a frequency value of 284 as an initial component of the bipartite cipher equivalents of plaintext letters, and a frequency value of only 169 as a final component.
Similarly, the letters V and X have frequency values of 47 and 69, respectively, as initial components and 120 and 198 as final components. It is obvious, then, that in this checkerboard the weighted frequency values of the letters A, V, and X as initial components differ considerably from the values of these same letters as final components, the value for G as an initial is only a little less than its value as a final, the values of D and F as initials are only a little more than their values as finals. But it is the wide variations in the weighted frequency values of certain of the letters as initial components and as final components, exemplified in the case of A, V, and X, which form the basis of the general solution, because these wide variations afford a means for making the various differentiations noted in subparagraph *d*.

f. Of course, in working with an unknown example, the composition of the checkerboard is unknown and therefore no accurate frequency weights may be assigned to the ADFGVX components in the cryptograms. However, it is still possible to arrive at some approximations for these weights in case there are several cryptograms available for study, as would normally be true in actual practice. How this can be done will be shown very soon, by studying an example. For the purposes of this study the set of 12 cryptograms given below will be used.

I

V D D G G G V F D F V D V V F V D G A D D A F F F
V D X F D D X D V X A D V D V F X G D F V A D D G
D G D G V G D D D F X F A D A V D V G D G A D X V
D A D A D F X A V F V D D A A V D F F D F V G D F
V D D G V D D D D A V A D A F A D D X A D D G A D
F V G F V D G A D V F X V X D G D D A G G D D X F
F D D X A D F G D A G X D D A V F D A F G V F V F
A F F V F A F X G F X D G V A D F V D G G A V G G
D D G D V X A X F D D X (212 letters)

II

V D A A V D D F X F X D D A X G X F X D D F X A D
V A G D D F A X D V A V D V D D F V F V F F G D G
F V A X V X A V G D V D X F D X D G A X G F G G F
V F G D F V D X A V X D D V G D D V G V A G F X F
A A A X D D X G (108 letters)

III

D A G A A F G A G V D A F G G X F D X D F V V X G
F X F D X D D A G A D D G V A D D V D D G A F G A
V G D G X D D D A V F V D D F D A A A A D X A G D
X A G G D D A V G V F G D V F V D G G X G G A F F
V F D A X G D D D G D A F D A D G G A D D G D X A
F V D F D X F V G D D V A V F D D D V F A G D F F
F X A A D F A D G G V F D A V D G X F V D A A V G
D X F G G D D X G D A (186 letters)

IV

A D X V F X V G G V F D D V A F G A A V F D G V D
D D G D G F D V V A F G X F X F D D D D V G D A X
D A X D D D A G V F F A A D V G D F X G X G V G D
D D D A D V X V F A V D A X X D F A A F A V D V G
V D V D D A X D A A (110 letters)

V

D F X F D D V V V D X F X F X F F F V A G F D X A
V D A G F D V D G F A D A A D F D V F G D A D F V
F V F X G X D D A G D V G V F D G X X D F F G D G
X G V D D V D D F G F V G D D V F V A G X X D F V
D X A V F G A G A G A X D V D F X G V G D A D D X
A G X D A D F D G X F D G G F V G X V V G D D D A
G X V D G V D V G X D D F D D V A G A A D G D D F
D G A G D F D D D D X G V G V G G G D G X D F G F
A D (202 letters)

VI

G D G F X A G V F V D D X G X D V D D A X D A A X
F A G V G D X F F V X F A D G F F D X A A F V X F
D F X F V G D G F X F D V V X V G D F V D D V F D
F V V D V D G G V F X F G V X F F V G V D D G D D
D D G D D A V G V X G A F F X F V D D D (120 letters)

VII

G A F G F F X F V F G F X A V A G G X D X X D D F
A G V D D V D V F F A D A V A V F V G G A D A A F
V F D F V D X F X X G D X D D F V D F F X D V F X
V A D X V A X D V X A F F V D F D G X F D G F D D
F V D V V A A F V F F V X D G F D D V A D D F D D
D X F F A G F X F X A A G V D G G V D F G G G X D
F D F V A F F G F X G D A X D G D G G D D A V D X
A D F A F V F X D D X V A G D V V D D F X D G X X
D V F V F D D D D A A F D F X D X G D A A F V D F
D V D D V A D D V D V A V D G A F V F X F A A V D
D F V D (254 letters)

VIII

D G V V G F X G G G A D F A F V V V A X A V G G V
V D V G V V D A V G D G D G A V F D D A D D D X X
D X F V F X G V G G D G D F G G D A D F D D X A V
F D D V F A D X G D A D G V A F F X A D F A D X D
G F A D F D D G V D V X A V A D D X F F A G D X F
F V F G F G F D F D V D X X D D G G D (144 letters)

IX

```
G D D D D   X V G V D   V D A V G   F G D F V   D V A V D
G F A G X   A V F F G   V A D D D   A X X A X   D G A D G
X A V V D   G X X A A   A V A D A   D G X D V   G D D D D
G V F X A   A V G G V   F X D A F   D G V G A   F G D D F
A V V G D   D V D F X   D V D G F   V A A G D   X F D V A
A D A G D   A X F V G   D D D A G   V A V F G   X X F D D
G X F V D   G G D A V   D A G G F   D A X D X   F F V G F
A X X A D   D F  (182 letters)
```

X

```
D G D D F   V F A V D   V F D A D   G F V G V   G G D F V
D V V X D   D F D D V   G X G V D   X G V G D   X D G D X
F X F D X   V D A A D   D F X D D   A F F A A   F V F A G
D A A G G   F A X G V   X X F X A   D G D F D   G X G D A
D A X G V   V V D A A   G G V F G   V A V F V   A A G A X
G X D G A  (130 letters)
```

XI

```
V F D D V   A X G D A   D F G G G   G F G D D   F X X D A
F D D X G   G A V G A   G D V D F   D F D D D   G A F A F
D A A A G   V A V F G   G V A D D   G D D F G   F V D D A
D F G A F   D F V D D   F V V V A   D A G D X   F X X X F
F D X G D   F D G F D   D F G D A   G F A A G   G A D X D
G V D G A   V G V D F   D D F X G   A G X F G   V F V V D
G V D X D   F F F X G   X G X A G   A G V G D   V V X G F
V D X D D   X F V D D   X  (186 letters)
```

XII

```
X F D F X   V V D V D   A V D A D   V F A G D   G V A D D
F D A A D   X A D F V   G V D G F   X F G D V   F V D D D
D G D V V   A V V V F   A D D A X   A V F V A   D A X D V
G D D F A   X D D G X   G V F X A   V X V F D   G D X D F
D V X A D   V A V A V   G V D D D   A F D F A   D V F F V
V G D A G   F X D D F   A D V X V   D F X F F   V V G F X
X G F V A   V F A G G   D A V V D   X D X G D   D V V A D
D D A G A   A G X F G   D D D G V   F G F V G   V X G V F
D F F D A   A D V D D   X G D F D   D V D D G   A F G D
(224 letters)
```

43. Illustration of solution.[7]—*a.* Since the initial letters of all 12 cryptograms are in the same class, that is, either initial or final components, they may all be combined into a single distribution. Furthermore, since it is certain that *regardless of whether the transposition rectangle has an odd or an even number of columns* the 3d, 5th, 7th . . . letters of the cryptograms are in the same class as the first letter, the 3d, 5th, 7th . . . letters may be added to the distribution, so long as these odd letters come from the same section (column 1). It is, however, necessary to limit the number of letters taken from the beginning of any one cryptogram to a reasonable length of column, depending on the size of the cryptogram. Assuming it is known that the enemy is using transposition keys of not less than 15 nor more than 22 numbers, the latter could be taken as the maximum possible size. But to be on the safe side it will be here assumed that a transposition rectangle of not more than 25 columns is being used. Hence, so far as concerns cryptogram I, which has 212 letters, on the basis of a key of 25 numbers [$(25\times9)-13=212$] there will be 12 columns of 9 letters and 13 columns of 8 letters. Since there is no way of telling which are long and which are short columns, it will be safer to work on the basis of columns of 8 letters. Therefore, the first 8 letters of cryptogram I are to be taken. In the case of cryptogram II, with 108 letters, its first 4 letters will be taken, and so on, through the 12 cryptograms, the number of letters to be taken in each case being governed by the length of the cryptogram. The sections taken in the case of the 12 cryptograms are shown in figure 65.

---

[7] This illustration uses the same cryptograms and follows quite closely along the lines employed in a technical paper of the Signal Intelligence Service entitled *General Solution for the ADFGVX Cipher*, prepared by Messrs. Rowlett, Kullback, and Sinkov, in 1934.

Cryptogram	Length	Letters taken	Cryptogram	Length	Letters taken
I	212	VDDGGGVF	VII	254	GAFGFFXFVF
II	108	VDAA	VIII	144	DGVVG
III	186	DAGAAFG	IX	182	GDDDDXV
IV	110	ADXV	X	130	DGDDF
V	202	DFXFDDVV	XI	186	VFDDVAX
VI	120	GDGF	XII	224	XFDFXVVD

FIGURE 65.

b. The odd and the even letters of these 12 sections are then distributed separately, the results being shown in figures 66 and 67. A consideration of the mechanics of this system leads to the expectation that *if the transposition rectangle has an even number of columns the two distributions will be similar; if it has an odd number, they will be different.* The similarity or difference between the two distributions is usually discernible with as few as 20 or 25 letters.

Odd (1st, 3d, . . .) letters

A	D	F	G	V	X
𝍷𝍷𝍷	𝍸 𝍸 𝍷	𝍷𝍷𝍷	𝍸 𝍷𝍷𝍷	𝍸 𝍸 𝍷	𝍸 𝍷

FIGURE 66.

Even (2d, 4th, . . .) letters

A	D	F	G	V	X
𝍸	𝍸 𝍸	𝍸 𝍸 𝍷	𝍸	𝍷𝍷𝍷𝍷	𝍷

FIGURE 67.

c. Letters V and X are of high frequency in the *odd* positions (fig. 66) but of low frequency in the *even* positions (fig. 67), whereas the letter F is of low frequency in the odd positions and of high frequency in the even positions. There can be no question that the two distributions are dissimilar, and the indications are clear that the transposition rectangle involves an odd number of columns.

d. Now the letters in figure 66 may be initial components, those in figure 67, final components, or the reverse may be the case. At the present stage of the study it is impossible to ascertain which of these alternative hypotheses is correct. However, this information is really immaterial at this stage. Suppose the letters in figure 66 are arbitrarily designated as class 1 components, those in figure 67 as class 2 components. Class 1 components (fig. 66) are characterized by a predominance of V's and X's (over their frequencies in fig. 67); class 2 components (fig. 67) are characterized by a predominance of F's (over its frequency in fig. 66).

e. The two distributions in figures 66 and 67 apply to the letters which come from column 1 of the transposition rectangles for the 12 cryptograms understudy. In this column, the V's and X's fall predominantly in the odd positions, the F's fall predominantly in the even positions. Therefore, beginning with position 1, the components in this column show an alternation of the type $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$. By referring to figure 63 it will become clear that if class 1 components are initial components, then it must follow that column 1 occupies an odd position in the transposition rectangle; but if class 1 components are final components, then it must follow that column 1 occupies an even position in the transposition rectangle. Which of these alternatives is true cannot be ascertained at the moment. *But the important point to be noted is that a definite reversal in the type of alternation of class 1 and class 2 components indicates the transit, in the transposition, from the end of one column to the beginning of the next column.* That is, if it is found that from the beginning of the cryptogram the alternation of components is $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$ and after a number of letters this alternation changes to $\Theta_2 \rightarrow \Theta_1 \rightarrow \Theta_2$, the point where this change occurs marks the end of column 1 and the beginning of the column 2. For the sake of brevity in reference, in the subsequent paragraphs the type of alternation $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$ will be designated as the "+ type," and this type of alternation characterizes columns which fall in the odd positions in the transposition rectangle *i. e.*, in the 1st, 3d, 5th, 7th, . . . positions from the left. The other type, $\Theta_2 \rightarrow \Theta_1 \rightarrow \Theta_2$ will be designated as the "— type," and this type of alternation characterizes columns which fall in the even positions in the transposition rectangle *i. e.*, in the 2d, 4th, 6th, 8th, . . . positions from the left.

f. With these principles in mind, let cryptograms III and XI, each containing 186 letters, be studied. They may be superimposed, since they have identical numbers of letters and therefore the columns end at exactly the same points in both cryptograms.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
III.	D	A	G	A	A	F	G	A	G	V	D	A	F	G	G	X	F	D	X	D	F	V	V
XI.	V	F	D	D	V	A	X	G	D	A	D	F	G	G	G	G	F	G	D	D	F	X	X

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
III.	X	G	F	X	F	D	X	D	D	A	G	A	D	D	G	V	A	D	D	V	D	D	G
XI.	D	A	F	D	D	X	G	G	A	V	G	A	G	D	V	D	F	D	F	D	D	D	G

	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
III.	A	F	G	A	V	G	D	G	X	D	D	D	A	V	F	V	D	D	F	D	A	A	A
XI.	A	F	A	F	D	A	A	A	G	V	A	V	F	G	G	V	A	D	D	G	D	D	F

	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
III.	A	D	X	A	G	D	X	A	G	G	D	D	A	V	G	V	F	G	D	V	F	V	D
XI.	G	F	V	D	D	A	D	F	G	A	F	D	F	V	D	D	F	V	V	V	A	D	A

	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115
III.	G	G	X	G	G	A	F	F	V	F	D	A	X	G	D	D	D	G	D	A	F	D	A
XI.	G	D	X	F	X	X	X	F	F	D	X	G	D	F	D	G	F	D	D	F	G	D	A

	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138
III.	D	G	G	A	D	D	G	D	X	A	F	V	D	F	D	X	F	V	G	D	D	V	A
XI.	G	F	A	A	G	G	A	D	X	D	G	V	D	G	A	V	G	V	D	F	D	D	F

	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161
III.	V	F	D	D	D	V	F	A	G	D	F	F	F	X	A	A	D	F	A	D	G	G	V
XI.	X	G	A	G	X	F	G	V	F	V	V	D	G	V	D	X	D	F	F	F	X	G	X

	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184
III.	F	D	A	V	D	G	X	F	V	D	A	A	V	G	D	X	F	G	G	D	D	X	G
XI.	G	X	A	G	A	G	V	G	D	V	V	X	G	F	V	D	X	D	D	X	F	V	D

	185	186
III.	D	A
XI.	D	X

FIGURE 68.

g. It has already been noted that beginning with the first letter of any one of the cryptograms, the type of alternation for column 1 is +. It is therefore not astonishing to find, within the first 10 letters, an alternation of the + type. Note how the V's and X's fall in the odd positions, the F's in the even. Thus:

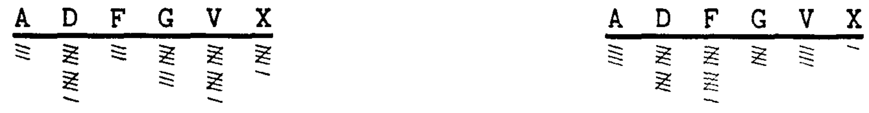
	1	2	3	4	5	6	7	8	9	10
III.	D	A	G	A	A	F	G	A	G	V
XI.	V	F	D	D	V	A	X	G	D	A

It is seen that there are 2 V's which fall in odd positions (1 and 5), but one V falls in an even position (10). There is an X, which falls in an odd position (7); there are 2 F's which fall in even positions (2 and 6). Unquestionably, then, the type of alternation, at least for the first 10 letters in each of these cryptograms, is +.

h. Take the next section of 10 letters in these two cryptograms. The letters are as follows:

	11	12	13	14	15	16	17	18	19	20
III.	D	A	F	G	G	X	F	D	X	D
XI.	D	F	G	G	G	G	F	G	D	D

Here there are 4 F's; 3 of them fall in odd positions (13, 17, 17), and one falls in an even position (12). There are 2 X's; one falls in an odd position (19), one in an even position (16). There are no V's among these letters. So far as the evidence afforded by the F's is concerned, it would appear that this section of text shows the type 2 or "— type" of alternation of components, since in type 1 or "+ type" the F's occupy even positions and here the majority of them occupy odd positions. But so far as the X's are concerned, the evidence is equally balanced: one X falls in an odd position, one in an even position. There being no V's, no conclusions can be drawn from this letter. To be guided solely by the evidence afforded by the 3 F's may be unwarranted. Is it not possible to weight the frequencies of the letters so that it will be unnecessary to rely merely upon a few of them and the evidence afforded by all the letters can be taken into account? Why not assign frequency weights according to the two distributions in figures 66 and 67? The figures then become as follows:



Odd (1st, 3d, . . .) letters

Total = 3 + 11 + 3 + 8 + 11 + 6 = 42 letters

Even (2d, 4th, . . .) letters

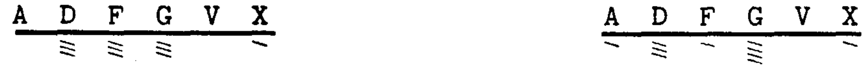
Total = 4 + 10 + 11 + 5 + 4 + 1 = 35 letters

FIGURE 69.

Since the odd letters have a total frequency of 42, the even, a total frequency of 35, for purposes of equalizing the distributions in applying the weights it seems advisable to deduct one-sixth from the total when applying the weights to odd letters.

i. Now in applying these weights to the letters, it must be borne in mind that since a transposition rectangle with an odd number of columns is involved, half of the letters are class 1 components, the other half are class 2 components. Hence, in finding the frequency value of the letters it is necessary to apply the weighted frequencies to alternate letters in the sections, as shown in figure 70.

	11	12	13	14	15	16	17	18	19	20
III.	D	A	F	G	G	X	F	D	X	D
XI.	D	F	G	G	G	G	F	G	D	D



Distribution of odd letters

f = 0 3 3 3 0

Distribution of even letters

f = 1 3 1 4 0 1

FIGURE 70.

These distributions, when evaluated in accordance with figure 69, yield a total frequency value of 126; when evaluated in accordance with figure 69 *reversed*, yield a total frequency value of 143. The detailed calculations are as follows:

On the basis of figure 69 normal (odd letters as θ_1's, even letters as θ_2's):

$$0(3)+3(11)+3(3)+3(8)+0(11)+1(6)=72$$

$$72-\frac{72}{6}=60$$

$$1(4)+3(10)+1(11)+4(5)+0(4)+1(1)=66 \qquad =66$$

$$\text{Total}=126$$

On the basis of figure 69 reversed (even letters as θ_1's, odd letters as θ_2's):

$$1(3)+3(11)+1(3)+4(8)+0(11)+1(6)=77$$

$$77-\frac{77}{6}=64$$

$$0(4)+3(10)+3(11)+3(5)+0(4)+1(1)=79 \qquad =79$$

$$\text{Total}=143$$

j. Now the frequency sums here obtained (126 *vs.* 143) indicate that an alternation of the type $\theta_2 \rightarrow \theta_1 \rightarrow \theta_2$ is in effect, that is, if a beginning is made with position 11, the type of alternation is "—". Since the type of alternation for the first 10 letters is "+" and for the second 10 letters "—", the reversal in alternation would indicate that column 1 of the transposition rectangle ends somewhere near the 10th letter. This same sort of reversal takes place after the 20th letter, as shown by the calculation in figure 71.

	21	22	23	24	25	26	27	28	29	30
III.	F	V	V	X	G	F	X	F	D	X
XI.	F	X	X	D	A	F	D	D	X	G

Distribution of odd letters

	A	D	F	G	V	X
	⟍	⟍⟍	⟍⟍	⟍	⟍	⟍⟍⟍
f=	1	2	2	1	1	3

Distribution of even letters

	A	D	F	G	V	X
		⟍⟍	⟍⟍⟍	⟍	⟍	⟍⟍⟍
f=	0	2	3	1	1	3

On the basis of figure 69 normal (odd letters as θ_1's, even letters as θ_2's):

$$1(3)+2(11)+2(3)+1(8)+1(11)+3(6)=68$$

$$68-\frac{68}{6}=57$$

$$0(4)+2(10)+3(11)+1(5)+1(4)+3(1)=65 \qquad =65$$

$$\text{Total}=122$$

On the basis of figure 69 reversed (even letters as θ_1's, odd letters as θ_2's):

$$0(3)+2(11)+3(3)+1(8)+1(11)+3(6)=68$$

$$68-\frac{68}{6}=57$$

$$1(4)+2(10)+2(11)+1(5)+1(4)+3(1)=58 \qquad =58$$

$$\text{Total}=115$$

FIGURE 71.

Beginning with the 21st position, the alternation is of type $\theta_1 \to \theta_2 \to \theta_1$; hence it is of the "+" type. Again the reversal in type of alternation occurs in passing from the 2d set of 10 letters to the 3d set, and this indicates that column 2 of the transposition rectangle ends somewhere near the 20th letter. But, fortunately, this time the exact location of the break is definitely indicated: The simultaneous appearance of V and X in the sequent positions 22 and 23 leads to the idea that the 22d letter marks the end of column 2 and the 23d letter marks the beginning of column 3. *There is nothing of an absolute nature in this point:* It is merely an indication based upon probabilities and does not constitute a conclusive proof by any means. Now if there is this definite break at the end of 22 letters it means that columns 1 and 2 must each contain 11 letters. The calculations have heretofore been based upon sections of 10 letters and the results are therefore modified as shown in the following calculation:

FIRST SECTION (letters 1–11)

	1	2	3	4	5	6	7	8	9	10	11
III.	D	A	G	A	A	F	G	A	G	V	D
XI.	V	F	D	D	V	A	X	G	D	A	D

Distribution of odd letters

	A	D	F	G	V	X
	/	/////		///	//	/
f=	1	5	0	3	2	1

Distribution of even letters

	A	D	F	G	V	X
	/////	/	//	/	/	
f=	5	1	2	1	1	0

Weighted values of distributions:

On the basis of figure 69 normal (odd letters as θ_1's, even letters as θ_2's):

$$1(3)+5(11)+0(3)+3(8)+2(11)+1(6)=110 \qquad 110-\frac{110}{6}=92$$

$$5(4)+1(10)+2(11)+1(5)+1(4)+0(1)=61 \qquad =61$$

$$\text{Total}=153$$

On the basis of figure 69 reversed (even letters as θ_1's, odd letters as θ_2's):

$$5(3)+1(11)+2(3)+1(8)+1(11)+0(6)=51 \qquad 51-\frac{51}{6}=42$$

$$1(4)+5(10)+0(11)+3(5)+2(4)+1(1)=78 \qquad =78$$

$$\text{Total}=120$$

The type of alternation is $\theta_1 \to \theta_2 \to \theta_1$, or "+".

SECOND SECTION (letters 12–22)

	12	13	14	15	16	17	18	19	20	21	22
III.	A	F	G	G	X	F	D	X	D	F	V
XI.	F	G	G	G	G	F	G	D	D	F	X

Distribution of odd letters

	A	D	F	G	V	X
		/	/////	///		
f=	0	1	5	3	0	1

Distribution of even letters

	A	D	F	G	V	X
	/	///	/	////	/	//
f=	1	3	1	4	1	2

Weighted values of distributions:

On the basis of figure 69 normal (odd letters as θ_1's, even letters as θ_2's):

$$0(3)+1(11)+5\ (3)+3(8)+0(11)+1(6)=56 \qquad 56-\frac{56}{6}=\ 47$$

$$1(4)+3(10)+1(11)+4(5)+\ 1(4)+2(1)=71 \qquad =71$$

$$\text{Total}=118$$

On the basis of figure 69 reversed (even letters as θ_1's, odd letters as θ_2's):

$$1(3)+3(11)+\ 1(3)+4(8)+1(11)+2(6)=94 \qquad 94-\frac{94}{6}=\ 78$$

$$0(4)+1(10)+5(11)+3(5)+\ 0(4)+1(1)=81 \qquad =81$$

$$\text{Total}=159$$

Since the distribution here begins with an even-numbered position (12), and the greatest total is obtained on the basis of figure 69 *reversed*, the type of alternation for the second section of 11 letters is therefore again $\theta_1 \rightarrow \theta_2 \rightarrow \theta_1$, or "+".

THIRD SECTION (letters 23–33)

	23	24	25	26	27	28	29	30	31	32	33
III.	V	X	G	F	X	F	D	X	D	D	A
XI.	X	D	A	F	D	D	X	G	G	A	V

Distribution of odd letters

	A	D	F	G	V	X
	//	///		//	//	///
f=	2	3	0	2	2	3

Distribution of even letters

	A	D	F	G	V	X
	/	///	///	/		//
f=	1	3	3	1	0	2

Weighted values of distributions:

On the basis of figure 69 normal (odd letters as θ_1's, even letters as θ_2's):

$$2(3)+3(11)+\ 0(3)+2(8)+2(11)+3(6)=95 \qquad 95-\frac{95}{6}=\ 79$$

$$1(4)+3(10)+3(11)+1(5)+\ 0(4)+2(1)=74 \qquad =\ 74$$

$$\text{Total}=153$$

On the basis of figure 69 reversed (even letters as θ_1's, odd letters as θ_2's):

$$1(3)+3(11)+\ 3(3)+1(8)+0(11)+2(6)=65 \qquad 65-\frac{65}{6}=\ 54$$

$$2(4)+3(10)+0(11)+2(5)+\ 2(4)+3(1)=59 \qquad =\ 59$$

$$\text{Total}=113$$

Since the best values are obtained on the basis of figure 69 normal, the type of alternation for the third section of 11 letters is $\theta_1 \rightarrow \theta_2 \rightarrow \theta_1$, or "+".

k. Now if columns 1 and 2 contain 11 letters, and the total number of letters is 186, the transposition rectangle obviously has 17 columns, there being 16 long columns of 11 letters and one short column of 10 letters [(17×11)−1=186].

l. There is another cryptogram which also contains but one short column, viz, VII, of 254 letters [17×15)−1=254]. The columns of this cryptogram contain 4 more letters than the corresponding columns of III and XI. Assuming, momentarily, that the last column is the short one, cryptogram VII may be added to the superposition of III and XI, provided these sets of 4 additional letters are accounted for. This has been done in figure 72. In that figure the 4 extra letters pertaining to cryptogram VII are shown as falling under the last letters of the columns of cryptograms III and XI, but this is only an arbitrary placement. It is sufficient to place these extra letters in such positions as will make the first one of the series begin in an even position.

m. Since the transposition rectangle is now known to be 17 columns wide, the data in figure 69 may be enlarged to correspond to this information. For example, whereas in originally constructing figure 69 the first column of cryptogram I was assumed to have only 8 letters (to correspond to a key of 25 numbers), it may now be extended to a column of 12 letters, and so on. The additional portions used to make the distributions in figure 74 are shown underlined in figure 73.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
III.	D	A	G	A	A	F	G	A	G	V	D	A	F	G	G	X	F	D	X	D	F	V
XI.	V	F	D	D	V	A	X	G	D	A	D	F	G	G	G	G	F	G	D	D	F	X
VII.	G	A	F	G	F	F	X	F	V	F	G	A	G	G	X	D	X	X	D	D	F	A
								F	X	A	V								G	V	D	D

	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
III.	V	X	G	F	X	F	D	X	D	D	A	G	A	D	D	G	V	A	D	D	V	D
XI.	X	D	A	F	D	D	X	G	G	A	V	G	A	G	D	V	D	F	D	F	D	D
VII.	V	D	V	F	F	A	D	A	V	A	V	A	D	A	A	F	V	F	D	F	V	D
								F	V	G	G								X	F	X	X

	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
III.	D	G	A	F	G	A	V	G	D	G	X	D	D	D	A	V	F	V	D	D	F	D
XI.	D	G	A	F	A	F	D	A	A	A	G	V	A	V	F	G	G	V	A	D	D	G
VII.	G	D	X	D	D	F	V	D	F	F	X	V	A	D	X	V	A	X	D	V	X	A
								D	V	F	X								F	F	V	D

	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88
III.	A	A	A	A	D	X	A	G	D	X	A	G	G	D	D	A	V	G	V	F	G	D
XI.	D	D	F	G	F	V	D	D	A	D	F	G	A	F	D	F	V	D	D	F	V	V
VII.	F	D	G	X	F	D	G	F	D	D	F	A	A	F	V	F	F	V	X	D	G	F
								V	D	V	V								D	D	V	A

	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110
III.	V	F	V	D	G	G	X	G	G	A	F	F	V	F	D	A	X	G	D	D	D	G
XI.	V	A	D	A	G	D	X	F	X	X	X	F	F	D	X	G	D	F	D	G	F	D
VII.	D	D	F	D	D	D	X	F	F	A	G	A	A	G	V	D	G	G	V	D	F	G
								F	X	F	X								G	G	X	D

	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132
III.	D	A	F	D	A	D	G	G	A	D	D	G	D	X	A	F	V	D	F	D	X	F
XI.	D	F	G	D	A	G	F	A	A	G	G	A	D	X	D	G	V	D	G	A	V	G
VII.	F	D	F	V	A	F	F	G	F	X	G	G	D	G	G	D	D	A	V	D	X	A
								D	A	X	D								D	F	A	F

FIGURE 72.

	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154
III.	V	G	D	D	V	A	V	F	D	D	D	V	F	A	G	D	F	F	F	X	A	A
XI.	V	D	F	D	D	F	X	G	A	G	X	F	G	V	F	V	V	D	G	V	D	X
VII.	V	F	X	D	D	X	V	A	G	D	V	X	D	G	X	X	D	V	F	V	F	D
								V	D	D	F								D	D	D	A

	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176
III.	D	F	A	D	G	G	V	F	D	A	V	D	G	X	F	V	D	A	A	V	G	D
XI.	D	F	F	F	X	G	X	G	X	A	G	A	G	V	G	D	V	V	X	G	F	V
VII.	A	F	D	F	X	D	X	G	D	A	A	D	V	D	D	V	A	D	D	V	D	V
								F	V	D	F								A	V	D	G

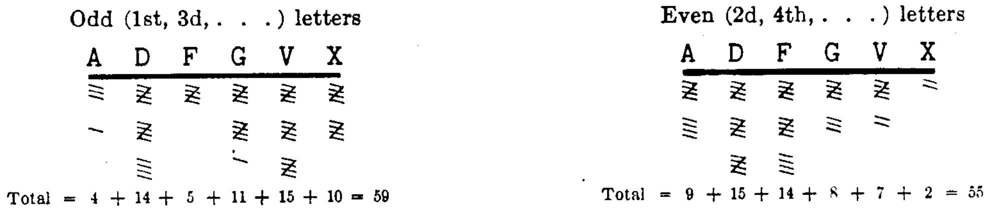
	177	178	179	180	181	182	183	184	185	186
III.	X	F	G	G	D	D	X	G	D	A
XI.	D	X	D	D	X	F	V	D	D	X
VII.	A	F	V	F	X	F	A	A	V	D
							D	F	V	D

FIGURE 72.—Continued.

Cryptogram	Length	Letters taken	Cryptogram	Length	Letters taken
I	212	VDDGGGVFDFVD	VII	254	GAFGFFXFVFGFXA
II	108	VDAAVD	VIII	144	DGVVGFXG
III	186	DAGAAFGAGV	IX	182	GDDDDXVGVD
IV	110	ADXVFX	X	130	DGDDFVF
V	202	DFXFDDVVVDX	XI	186	VFDDVAXGDA
VI	120	GDGFXAG	XII	224	XFDFXVVDVDAVD

FIGURE 73.

The new frequency weights are therefore as follows:



Total = 4 + 14 + 5 + 11 + 15 + 10 = 59

Total = 9 + 15 + 14 + 8 + 7 + 2 = 55

FIGURE 74.

Since the two totals are quite close together, no correction need be made of the nature of that made in preceding calculations, where one-sixth was deducted from the total values of odd letters.

n. Beginning with position 23, in the case of cryptograms III and XI, the next 11 letters, and, in the case of cryptogram VII, the next 15 letters are clearly of the "+" type of alternation. The data are as follows:

	23	24	25	26	27	28	29	30	31	32	33
III.	V	X	G	F	X	F	D	X	D	D	A
XI.	X	D	A	F	D	D	X	G	G	A	V
VII.	V	D	V	F	F	A	D	A	V	A	V
								F	V	G	G

Distribution of odd letters

	A	D	F	G	V	X
	//	////	/	///	///// //	///
f=	2	4	1	3	7	3

Distribution of even letters

	A	D	F	G	V	X
	////	////	/////	//		//
f=	4	4	5	2	0	2

Weighted values of distributions:

On the basis of figure 74 normal (odd letters as Θ_1's, even letters as Θ_2's):

$$
\begin{aligned}
2(4)+4(14)+1(5)+3(11)+7(15)+3(10)&=237\\
4(9)+4(15)+5(14)+2(8)+0(7)+2(2)&=\overline{186}\\
\text{Total}&=423
\end{aligned}
$$

On the basis of figure 74 reversed (even letters as Θ_1's, odd letters as Θ'_2's):

$$
\begin{aligned}
4(4)+4(14)+5(5)+2(11)+0(15)+2(10)&=139\\
2(9)+4(15)+1(14)+3(8)+7(7)+3(2)&=\overline{171}\\
\text{Total}&=310
\end{aligned}
$$

Since the greatest total is obtained on the basis of figure 74 normal, the type of alternation for the third section of letters is $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$, or "+".

o. Continuing the foregoing process with the letters beyond position 33, the data are as follows:

	34	35	36	37	38	39	40	41	42	43	44
III.	G	A	D	D	G	V	A	D	D	V	D
XI.	G	A	G	D	V	D	F	D	F	D	D
VII.	A	D	A	A	F	V	F	D	V	F	D
								X	F	X	X

Distribution of odd letters

	A	D	F	G	V	X
	///	///// ///	/		///	//
f=	3	8	1	0	3	2

Distribution of even letters

	A	D	F	G	V	X
	///	/////	/////	////	//	/
f=	3	5	5	4	2	1

Weighted values of distributions:

On the basis of figure 74 normal (odd letters as Θ_1's, even letters as Θ_2's):

$$
\begin{aligned}
3(4)+8(14)+1(5)+0(11)+3(15)+2(10)&=194\\
3(9)+5(15)+5(14)+4(8)+2(7)+1(2)&=\overline{220}\\
\text{Total}&=414
\end{aligned}
$$

On the basis of figure 74 reversed (even letters as Θ_1's, odd letters as Θ_2's):

$$
\begin{aligned}
3(4)+5(14)+5(5)+4(11)+2(15)+1(10)&=191\\
3(9)+8(15)+1(14)+0(8)+3(7)+2(2)&=\overline{186}\\
\text{Total}&=377
\end{aligned}
$$

Since the distribution begins here with an even-numbered position (34), and the greatest total is obtained on the basis of figure 74 normal, hence the alternation for the fourth section or column is of the type $\theta_2 \rightarrow \theta_1 \rightarrow \theta_2$, or "—".

p. (1) The data for the letters beyond position 44 are as follows:

	45	46	47	48	49	50	51	52	53	54	55
III.	D	G	A	F	G	A	V	G	D	G	X
XI.	D	G	A	F	A	F	D	A	A	A	G
VII.	G	D	X	D	D	F	V	D	F	F	X
								D	V	F	X

Distribution of odd letters

	A	D	F	G	V	X
	////	𝍸	/	///	///	////
f=	4	5	1	3	3	4

Distribution of even letters

	A	D	F	G	V	X
	///	////	𝍸 /	////		
f=	3	4	6	4	0	0

Weighted values of distributions:

On the basis of figure 74 normal (odd letters as θ_1's, even letters as θ_2's):

$$\begin{aligned} 4(4)+5(14)+1(5)+3(11)+3(15)+4(10)&=209\\ 3(9)+4(15)+6(14)+4(8)+0(7)+0(2)&=\underline{203}\\ \text{Total}&=412 \end{aligned}$$

On the basis of figure 74 reversed (even letters as θ_1's, odd letters as θ_2's):

$$\begin{aligned} 3(4)+4(14)+6(5)+4(11)+0(15)+0(10)&=142\\ 4(9)+5(15)+1(14)+3(8)+3(7)+4(2)&=\underline{178}\\ \text{Total}&=320 \end{aligned}$$

Since the distribution starts with an odd position (45) and the greatest total is obtained on the basis of figure 74 normal, the type of alternation for the fifth section or column is $\theta_1 \rightarrow \theta_2 \rightarrow \theta_1$, or "+".

q. The types of alternation for the first 5 columns, which are all long columns, is therefore + + + — +. Since cryptograms III and XI contain but one short column, it is advisable to be on the lookout for it as the work progresses. It is possible to continue with the process detailed above. For example, the calculations for the next or sixth section of 11 letters are shown below:

	56	57	58	59	60	61	62	63	64	65	66
III.	D	D	D	A	V	F	V	D	D	F	D
XI.	V	A	V	F	G	G	V	A	D	D	G
VII.	V	A	D	X	V	A	X	D	V	X	A
								F	F	V	D

Distribution of odd letters

	A	D	F	G	V	X
	𝍸	////	////	/	/	//
f=	5	4	4	1	1	2

Distribution of even letters

	A	D	F	G	V	X
	/	𝍸 //	/	//	𝍸 ///	/
f=	1	7	1	2	8	1

Weighted values of distributions:

On the basis of figure 74 normal (odd letters as Θ_1's, even letters as Θ_2's):

$$\begin{array}{r} 5(4)+4(14)+\ 4(5)+1(11)+1(15)+2(10)=142 \\ 1(9)+7(15)+1(14)+\ 2(8)+\ 8(7)+1(2)\ =\underline{202} \\ \text{Total}=344 \end{array}$$

On the basis of figure 74 reversed (even letters as Θ_1's, odd letters as Θ_2's):

$$\begin{array}{r} 1(4)+7(14)+\ 1(5)+2(11)+8(15)+1(10)=259 \\ 5(9)+4(15)+4(14)+\ 1(8)+\ 1(7)+\ 2(2)=\underline{180} \\ \text{Total}=439 \end{array}$$

Since the distribution starts with an even position (56) and the greatest total is obtained on the basis of figure 74 reversed, the type of alternation for the sixth section or column is $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$, or "+".

r. But perhaps advantage should be taken of the availability of additional cryptograms. For example, cryptogram V, of 202 letters, has 2 short columns [(17×12)−2=202], whereas the cryptograms thus far dealt with each have but one. That is, cryptogram V has one short column in common with cryptograms III, XI, and VII, and one additional short column not possessed by the latter. Can this additional short column of cryptogram V be located?

s. Suppose column 1 of cryptogram V is the additional short column. Then the letters of column 2 would be F X F X F F F V A G F D. These letters when evaluated on the basis of figure 74 normal yield a total of 77; when weighted on the basis of figure 74 reversed, a total of 144. The calculation is as follows:

Distribution of odd letters

	A	D	F	G	V	X
	/		/////			
f=	1	0	5	0	0	0

Distribution of even letters

	A	D	F	G	V	X
		/	/	/	/	//
f=	0	1	1	1	1	2

On the basis of figure 74 normal (odd letters as Θ_1's, even letters as Θ_2's):

$$\begin{array}{r} 1(4)+0(14)+\ 5(5)+0(11)+0(15)+0(10)=29 \\ 0(9)+1(15)+1(14)+\ 1(8)+\ 1(7)+\ 2(2)=\underline{48} \\ \text{Total}=77 \end{array}$$

On the basis of figure 74 reversed (even letters as Θ_1's, odd letters as Θ_2's):

$$\begin{array}{r} 1(9)+0(15)+5(14)+\ 0(8)+\ 0(7)+\ 0(2)=79 \\ 0(4)+1(14)+\ 1(5)+1(11)+1(15)+2(10)=\underline{65} \\ \text{Total}=144 \end{array}$$

According to this calculation, column 2 of cryptogram V seems to correspond to the type of alternation $\Theta_2 \rightarrow \Theta_1 \rightarrow \Theta_2$, that is "−". But from previous work it is fairly certain that column 2 is of the "+" type. Hence, column 1 of cryptogram V is probably not the additional short column of that message. Assuming column 2 to be the extra short column, no such contradiction is obtained, for the calculation is as follows:

Assuming column 2 to be short, the letters of column 3 are X A V D A G F D V D G F.

Distribution of odd letters

	A	D	F	G	V	X
	—		—	—	=	—
f=	1	0	1	1	2	1

Distribution of even letters

	A	D	F	G	V	X
	—	≡	=	—		
f=	1	3	1	1	0	0

Weighted values of distributions:

On the basis of figure 74 normal (odd letters as Θ_1's, even letters as Θ_2's):

$$\begin{array}{r} 1(4)+0(14)+1(5)+1(11)+2(15)+1(10) = 60 \\ 1(9)+\quad 3(15)+1(14)+1(8)+90(7)+0(2) = \underline{76} \\ \text{Total} = 136 \end{array}$$

On the basis of figures 74 reversed (even letters as Θ_1's, odd letters as Θ_2'2):

$$\begin{array}{r} 1(9)+0(15)+1(14)+1(8)+2(7)+1\ (2) = 47 \\ 1(4)+3(14)+1(5)+1(11)+0(15)+0(10) = \underline{62} \\ \text{Total} = 109 \end{array}$$

Since the greatest total is obtained on the basis of figure 74 normal, the type of alternation is $\Theta_1 \rightarrow \Theta_2 \rightarrow \Theta_1$ and column 3 is a "+" column, which is consistent with the formula + + + − + for columns 1 to 5, as previously ascertained.

If all the foregoing reasoning is correct, and column 2 is the additional short column for cryptogram V, it must be the next to the last column of the transposition rectangle. Since it is a "+" column, the last column must be a "−" one; therefore, there are 9 "−" columns and 8 "+" columns. This definitely determines that the "−" columns are the odd ones, the "+" columns the even ones, since in an odd-width rectangle there is one more odd column than even columns.

t. The single short column which is common to cryptograms III, XI, and VII is one of the columns beyond column 5. Assuming each possibility in turn, there is obtained for the type of alternation in each column the distributions of "+" and "−" shown in figure 75.

Assumption	Column																	Summation of +'s and −'s
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
(1) 6th short	+	+	+	−	+	+	+	+	−	+	+	+	−	−	−	−	−	10+, 7−
(2) 7th short	+	+	+	−	+	+	−	+	−	+	+	+	−	−	−	−	−	9+, 8−
(3) 8th short	+	+	+	−	+	+	−	−	−	+	+	+	−	−	−	−	−	8+, 9−
(4) 9th short	+	+	+	−	+	+	−	−	+	+	+	+	−	−	−	−	−	9+, 8−
(5) 10th short	+	+	+	−	+	+	−	−	+	−	+	+	−	−	−	−	−	8+, 9−
(6) 11th short	+	+	+	−	+	+	−	−	+	−	−	+	−	−	−	−	−	7+, 10−
(7) 12th short	+	+	+	−	+	+	−	−	+	−	−	−	−	−	−	−	−	6+, 11−
(8) 13th short	+	+	+	−	+	+	−	−	+	−	−	−	+	−	−	−	−	7+, 10−
(9) 14th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	−	−	−	8+, 9−
(10) 15th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	+	−	−	9+, 8−
(11) 16th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	+	+	−	10+, 7−
(12) 17th short	+	+	+	−	+	+	−	−	+	−	−	−	+	+	+	+	+	11+, 6−

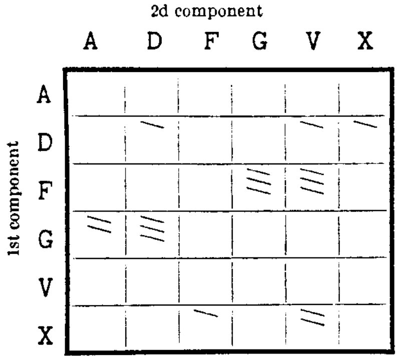
FIGURE 75.

u. The correct assumption must satisfy the following conditions:

(*a*) There must be 9 "−" and 8 "+" columns.

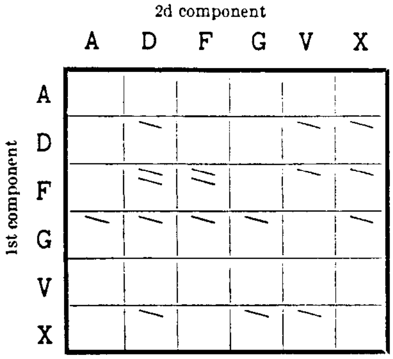
(*b*) The short column must be "−".

Only assumptions (3) and (5), in which column 8 and column 10 are short columns, satisfy these conditions. Therefore, column 2 is followed by either column 8 or 10. Testing the combination 2–8 for monoalphabeticity of bipartite pairs, the distribution shown in figure 76 is obtained. When combination 2–10 is tested, the distribution shown in figure 77 is obtained. Obviously, the 2–8 combination is the better.



E (ϕ) = .0667 × 17 × 16 = 18.14
ϕ = 22

FIGURE 76.



E (ϕ) = .0667 × 17 × 16 = 18.14
ϕ = 4

FIGURE 77.

v. It is possible by introducing cryptograms with additional short columns to determine more of the key. Thus, it was found by using cryptograms XII and VI that the first 3 numbers of the transposition key are 16–5–7. But the process of anagramming will yield the solution at least as rapidly. In this process, of course, advantage may be taken of the fact that the columns have been classified into the "+" and "−" types and no combinations of two "+" or two "−" columns need be tested, since only combinations of the type "+ −" or "− +" are permissible.

w. The final transposition key and the substitution checkerboard are shown in figure 78.

16	5	7	6	9	3	14	1	13	11	17	10	4	12	15	2	8
V	I	K	I	N	G	S	C	R	O	W	N	H	O	T	E	L

2d component

1st component	A	D	F	G	V	X
A	V	I	9	K	N	G
D	7	S	C	3	R	O
F	W	H	8	T	E	5
G	L	A	1	B	2	D
V	4	F	6	J	0	M
X	P	Q	U	X	Y	Z

FIGURE 78.

x. All the foregoing details concern a case in which the transposition rectangle has an odd number of columns. Now if the rectangle contains an even number of columns, this type of solution is, of course, still applicable, and in fact is easier, since the letters of the text of the respective sections do not have to be distributed into odd and even letters. It is only necessary to identify a section as being composed of initial components or of final components. This analysis then produces a series of sections corresponding in number with the number of columns in the transposition rectangle. This number will, of course, be even. By a careful study of where alternations in composition of components (Θ_1 or Θ_2) occur, the division of the text into sections corresponding to long and short columns can be accomplished. The remaining steps are obvious and follow the lines elucidated in paragraph 39*e–j*.

y. The entire structure upon which this general solution rests is destroyed if the substitution checkerboard has been consciously manipulated to equalize or flatten out the sums of the weighted frequencies of the letters in its rows and columns. For example, note the following checkerboard, which is not "perfect" but gives approximately similar frequencies in its rows and columns.[8]

		2d component						
		A	D	F	G	V	X	Sums
1st component	A	I 74	Q 3	S 61		U 26		164
	D		T 92		W 16	C 31	P 27	166
	F	G 16		A 74			N 79	169
	G	X 5	V 15	J 2	E 130	B 10	K 3	165
	V	R 76	M 25	F 28		L 36		165
	X		D 42	Z 1	Y 19	O 75	H 34	169
Sums		171	177	166	165	158	163	1,000

FIGURE 79.

z. If the statistical calculations upon which this general solution is based make use of the logarithms of the frequencies instead of the frequencies themselves, much more accurate and clear-cut data will be obtained.

---

[8] The frequencies indicated as those given in fig. 3, p. 13, *Military Cryptanalysis*, Part I.

SECTION X

SOLUTION OF BIFID FRACTIONATING SYSTEMS

Paragraph
Review of principles underlying the cryptographic method 44
Example of a simple bifid cipher 45
Principles of solution 46
Example of solution 47
Special solution 48
Periodic bifid ciphers 49
General principles underlying the solution 50
Ascertaining the length of the period 51
Illustration of solution 52
Special solutions for bifid systems 53
Solution of trifid systems 54
Concluding remarks on fractionating systems 55
Concluding remarks on transposition systems 56

44. Review of principles underlying the cryptographic method.—Several bifid fractionating systems have been explained in previous texts of this series.[1] In certain of these systems four basic steps are involved, two of substitution and two of transposition. These steps may be briefly described as follows: (1) A process of decomposition (substitution), in which each plain-text letter is replaced by two components, θ_c^1 and θ_c^2, of a bifid or bipartite alphabet; (2) a process of separation (transposition), in which the $\theta^1\theta_c^2$ components originally paired together are separated; (3) a process of recombination (transposition), in which the separated components are combined to form new pairs; (4) a process of recomposition (substitution), in which each new pair of components is given a letter value according to the original or a different bifid alphabet.

45. Example of a simple bifid cipher.—*a*. One of the simplest bifid fractionating systems is that exemplified in the following subparagraphs. It will be employed to set forth certain principles in the general solution of systems of this and similar nature.

b. Given the 25-cell substitution checkerboard shown in figure 80, let the message to be enciphered be ONE PLANE REPORTED LOST AT SEA. The first step is to replace the plain-text letters by the bipartite equivalents, the two elements or components being set down vertically beneath the plain-text letters. This represents the first two of the four processes referred to in paragraph 44, the first being that of decomposition or substitution, the second, that of separation or transposition, represented by the manner in which the two bipartite elements are set down vertically (instead of horizontally), thus separating the two elements from their normal horizontal juxtaposition.

---

[1] See Special Text No. 166, *Advanced Military Cryptography*, sec. XI and *Military Cryptanalysis, Part I*, sec. IX.

2d component

		1	2	3	4	5
1st component	1	M	A	N	U	F
	2	C	T	R	I	G
	3	B	D	E	H	K
	4	L	O	P	Q	S
	5	V	W	X	Y	Z

FIGURE 80.

```
Plain text......O N E   P L A N E   R E P O R T E D   L O S T   A T   S E A
Components..{4  1 3   4 4 1 1 3   2 3 4 4 2 2 3 3   4 4 4 2   1 2   4 3 1
            {2  3 3   3 1 2 3 3   3 3 3 2 3 2 3 2   1 2 5 2   2 2   5 3 2
```

The third process, that of recombination or recomposition, also involving a transposition, is now to be performed and will consist in combining elements standing in diagonal relationship to the right, that is, as shown by the arrows below:

```
O N E   P L A N E   R E P O R T E D   L O S T   A T   S E A
4 1 3   4 4 1 1 3   2 3 4 4 2 2 3 3   4 4 4 2   1 2   4 3 1
 ↗ ↗  ↗  ↗ ↗ ↗ ↗
2 3 3   3 1 2 3 3   3 3 3 2 3 2 3 2   1 2 5 2   2 2   5 3 2
```

giving the pairs 21, 33, 34, 34, 11, 21, 33, etc. There are left, at the end of the process, one element in the upper line at the extreme left and another element in the lower line at the extreme right, yielding the pair 24, which may be placed at the head or tail of the resultant combinations, as preagreed. The last or fourth process, that of recomposition or substitution, is to replace the new pairs of components by letters from the original or a new checkerboard. If the same checkerboard is used, it yields the text shown herewith:

```
Plain...........O N E   P L A N E   R E P O R T E D   L O S T   A T   S E A
Components..{4  1 3   4 4 1 1 3   2 3 4 4 2 2 3 3   4 4 4 2   1 2   4 3 1(4)
             ↗ ↗ ↗   ↗ ↗ ↗ ↗
            {2  3 3   3 1 2 3 3   3 3 3 2 3 2 3 2   1 2 5 2   2 2   5 3 2
Cipher..........C E H   H M C E D   E H H T D R E I   U I W C   T I   X·B I
```

c. Another and perhaps more simple way of accomplishing the same process is to set down the bipartite equivalents horizontally and recombine them as shown below:

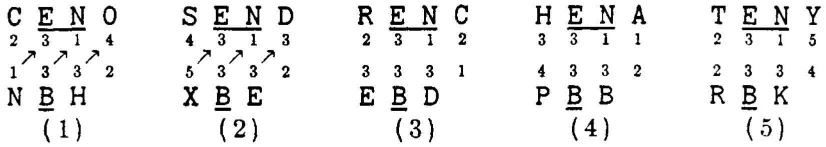
```
 O    N    E    P    L    A    N    E
42   13   33   43   41   12   13   33  . .
  ↖↗   ↖↗   ↖↗   ↖↗   ↖↗   ↖↗   ↖↗   ↖↗
   C    E    H    H    M    C    E
```

The results are identical with those obtained from the preceding manner of operation. The text is of course sent in 5-letter groups.

d. Instead of using the digits 1, 2, 3, 4, 5, as the bipartite components one can use the vowels A, E, I, O, U, or any other characters that are deemed suitable. Perhaps digits are best as they are less likely to be confused with letters of the text.

e. As intimated above, the checkerboard used for the recomposition may be different from that employed in the decomposition. But it will be shown that the additional safety afforded by using two different checkerboards is somewhat illusory, and is by no means as great as may appear on first consideration.

46. Principles of solution.—*a.* Note the following skeleton encipherments, using the checkerboard shown in figure 80:



These five encipherments have in common a plain-text digraph EN. The five cipher versions, however, have only a single letter in common, B. This is, of course, a phenomenon already encountered many times by the student and its cause is easily understood by him: The mechanics of the system tend to reduce by one character the lengths of the repetitions in the cipher text, as compared with their lengths in the plain text, a trigraphic repetition in the plain text manifesting itself as a digraphic repetition in the cipher text, a tetragraphic one becoming a trigraphic, and so on. More will later be stated on this phase of the matter.

b. But now study the individual cipher letter immediately preceding and succeeding the cipher letter which these five encipherments have in common. They are as follows:

Letters preceding B_c............N, X, E, P, R
Letters succeeding B_c............H, E, D, B, K

Reference to the checkerboard discloses the very interesting and important fact that the letters preceding the cipher repetition (B_c) *all come from the same column in the checkerboard, the letters succeeding the repetition all come from the same row in the checkerboard.* How this phenomenon is brought about is quite simple to see. Take the first of the five examples, that in which C E N O_p produces N B H_c. The N_c is the result of combining the second component of the bipartite equivalent of C_p with the first component bipartite equivalent of E_p, yielding the combination 13, which is N. No matter what the other three letters in the plain-text tetragraph may be, if the second letter is E_p, the second component bified equivalent of the first letter of the cipher trigraph must be a 3. This means that this first letter of the cipher trigraph must come from column 3 of the checkerboard. Exactly which row this letter will come from is determined by the identity of the second component of the bifid equivalent of the first letter of the plain-text tetragraph. Hence, since the 5 tetragraphs in the example all have the same plain-text letter in the second position, the initial letters of the cipher trigraphs all must come from the same column of the checkerboard. It is unnecessary to go through the reasoning, which is parallel, in the case of the third letters of the cipher trigraphs: these all must come from the same row of the checkerboard.

c. A good understanding of the phenomenon just noted can certainly be employed to advantage in solving this and similar types of systems, *for it becomes obvious that a careful study of the letters immediately preceding and following cipher repetitions should facilitate a reconstruction of the checkerboard employed in the substitution.*[1] Indeed, if there were no other phenomena to disturb this very simple relationship, solution would be quite easy. All that would be required would be to study the prefixes and suffixes to all the A's, B's, C's, . . . in the cryptogram, find the letters which belong in the same columns and rows of the checkerboard, and the reconstruction of the latter would follow very simply. Unfortunately, however, there is a disturbing phenomenon which must now be considered.

---

[1] The principle involved in such reconstruction was, to my knowledge, first pointed out and successfully employed early in 1938 by Associate Cryptanalysts S. Kullback and A. Sinkov.

d. Note the following encipherments:

Plain	L		P		U		R
	41		43		14		23
Cipher		U		B		O	
				(6)			

O		R		M		I
42		23		11		24
	T		B		A	
			(7)			

Here the B_c is preceded by letters (U and T) which not only are not in the same column as those in the corresponding position in the case of the first five encipherments, but also these two letters are themselves in different columns. The cause of this is not difficult to see. It is merely that the second component of the P_p and the second component of the R_p *happen* to be identical, the first component of the U_p and the first component of the M_p also *happen* to be identical, thus producing the same cipher letter in both cases. This is a phenomenon which must happen by chance a certain number of times, a number which is dependent not only upon the mechanics of the system but also upon the exact composition of the checkerboard. Disregarding for the moment the latter factor, it is obvious that if the checkerboard is perfectly balanced, the bifid element 3, for example, should occur 20 percent of the time as the first or as the second element of a bifid pair, since there are 5 elements and each can theoretically appear an equal number of times. However, since the checkerboard is not perfectly balanced, the bifid element 5 can, in the case of figure 80, appear as a second component of the bipartite equivalent of a cipher letter only very rarely, since it corresponds to the first component of the bifid equivalents of the letters V, W, X, Y, and Z, all of which are of low frequency. On the other hand, the bifid element 3, in the case of figure 80, can appear very frequently as a first component of the bifid equivalent of a cipher letter because it is the second component of the bifid equivalents for the high-frequency plain-text letters N, R, and E, which are all in column 3. However, since the exact composition of the checkerboard is unknown when cryptograms of this sort are to be solved, frequency weights can, of course, not be assigned to any of the components or bipartite elements and it will have to be assumed that each one has an equal probability of occurrence, that is, one-fifth.

e. From the foregoing discussion it is obvious that it would be unwise merely to study the prefixes and suffixes to identical single letters of the cipher text in an attempt to solve cryptograms of this sort, for the disturbing effect of the accidental identities of certain cipher letters would be sufficient to retard solution. A few detailed examples of the type of study that must be made in connection with repetitions in such systems as this will now be given.

f. It was stated in subparagraph *a* that the mechanics of the system tend to reduce by one character the lengths of the repetitions in the cipher text. The expression "tend to reduce" aptly describes the situation, for not only can it happen that a 3-letter repetition in the plain text may appear to remain a 3-letter repetition in the cipher text, but also it can happen that a 3-letter repetition in the plain text may even appear as a pseudo 4-letter repetition in the cipher text. Study the following examples (based on fig. 80) and note what happens in each case:

(*A*)	F		T		H		E		C
	15		22		34		33		21
		W		R		P		D	
(*B*)	Y		T		H		E		Y
	54		22		34		33		54
		O		R		P		K	

(1)

A 3-letter plain text repetition appears as a 2-letter cipher text repetition.

(*A*)	N		T		H		E		D		A
	13		22		34		33		32		12
		D		R		P		E		C	
(*B*)	T		T		H		E		H		A
	22		22		34		33		34		12
		T		R		P		E		L	

(2)

A 3-letter plain-text repetition appears as a 3-letter cipher-text repetition because the 1st components of the D_p and H_p happen to be identical (D and H are in same row in checkerboard).

```
(A)   I    N    T    H    E    G          (A)   T    O    T    H    E    M    A
      24   13   22   34   33   25               22   42   22   34   33   11   12
         L    D    R    P    D                     I    T    R    P    B    M

(B)   O    R    T    H    E    U          (B)   N    D    T    H    E    N    E
      42   23   22   34   33   14               13   32   22   34   33   13   33
         T    D    R    P    B                     E    T    R    P    B    E

              (3)                                         (4)
```

(3) A 3-letter plain-text repetition appears as a 3-letter cipher-text repetition because the second component of the N_p and R_p happen to be identical (R and N are in same column in checkerboard).

(4) A 3-letter plain-text repetition appears as a 4-letter cipher-text repetition because the phenomena of case 2 and case 3 occur simultaneously (O and D are in same column; M and N are in the same row in the checkerboard).

g. From a study of these phenomena the rule may be deduced that an n-letter repetition in the plain text is really reduced to an $(n-1)$-letter repetition in the cipher text, but it can happen fortuitously[3] that the real repetition is extended on either or both ends of the repetition by a pseudo-repetitious letter. Hence, a 3-letter plain-text repetition may appear as a 2-, 3-, or 4-letter repetition in the cipher-text.

h. It is therefore possible to make wholly erroneous deductions from some repetitions, especially if the latter are short. Note for instance the following example, still using Fig. 80 as a basis:

```
F    O    U    R    D    A   (YS)        S    O    M    E    D    U   (TY)
15   42   14   23   32   12   . . .      45   42   11   33   32   14   . . .
   Y    C    O    D    C                    Y    C    N    D    C
```

Here are 2 sequences of 5 cipher letters, identical save in the central letter, and yet the 6-letter plain text sequences have only 2 letters in common. This example is cited to show that the cryptanalyst must be very careful in respect to the deductions he may make in the case of short repetitions. In the example cited it happens that the accidental repetitions are such as to make the sequences as a whole almost appear to be identical.

i. It is these pseudo-repetitious elements which complicate the solution of what would otherwise be a simple system. To illustrate what is meant, note that in case (1) of subparagraph *f* the letters W_c and O_c, the prefixes to the repetition RP_c, do actually come from the same column of the checkerboard; the letters D_c and K_c, the suffixes, do actually come from the same row. But now note in case (2) that while the prefixes D_c and T_c come from the same column, the suffixes C_c and L_c do not come from the same row. Note also in case (3) that while the prefixes L_c and T_c do not come from the same column, the suffixes, D_c and B_c, do come from the same row; while in case (4) the prefixes turn out to be the same letter, T_c (which constitutes an example where the two prefixes come from the same column) but the suffixes, M_c and E_c, come from different rows. Since the exact length of the real repetition, without its pseudo-repetitious elements, does not readily manifest itself in the cipher text (although in favorable cases it may be deduced by a careful detailed analysis and comparison with nearly similar repetitions) the nature of the difficulties confronting the cryptanalyst become apparent.

j. The nature of the detailed analysis and comparison of repetitions referred to above may require a few words of explanation. Suppose that a cryptogram shows many occurrences of RP_c ($=THE_p$ in the foregoing examples). It would indicate a high-frequency plain-text trigraph. A few repetitions of such cipher trigraphs as RPE_c, DRP_c, $TRPB_c$, would lead to the surmise that the latter are of the type where pseudo-repetitious elements have crept into the picture and therefore the cryptanalyst should be very hesitant to assume that the adventitious prefixed letters are in the same columns, or that the adventitious suffixed letters are in the same row. In fact, he would be warranted in tentatively assuming the very opposite condition, that they are not in the same columns or rows, respectively. The conclusions derivable from a study of short repetitions can be carried over to the longer ones. Note the following four cases from which several conclusions may be reached:

---

[3] Strictly speaking, of course, not really fortuitously. It depends upon the exact letters which precede or follow the plain-text repetition and the exact positions these letters occupy in the checkerboard.

```
(1)  U   R   P   O   S   I   T   I   O   N   S   A
     14  23  43  42  45  24  22  24  42  13  45  12
       O   H   H   I   W   O   T   Q   C   H   V

(2)  I   S   P   O   S   I   T   I   O   N   O   F
     24  42  43  42  45  24  22  24  42  13  42  15
       Q   Y   H   I   W   O   T   Q   C   H   C

(3)  H   E   P   O   S   I   T   I   O   N   D   E
     34  33  43  42  45  24  22  24  42  13  32  33
       P   H   H   I   W   O   T   Q   C   E   R

(4)  O   R   P   O   S   I   T   I   O   N   L   U
     42  23  43  42  45  24  22  24  42  13  41  14
       T   H   H   I   W   O   T   Q   C   H   M
```

First, the 7-letter cipher sequence H I W O T Q C is common to all four cases; if only the cipher text were available, one could conclude that the plain-text repetition consists of 8 letters. Second, the letters H and Y probably come from the same column in the checkerboard, but as for O, P, and T, they may or may not come from the same column, most probably not. (Actually, O and T do, but P does not come from the same column as these 2 letters.) Third, the letters H and E probably come from the same row in the checkerboard, but as for V, C, and M, they may or may not come from the same row, most probably not. (Actually, all 3 letters come from different rows.)

k. Note the following cases of encipherment: The fact that the 7-letter cipher sequence is common to all four cases means that the plain-text repetition consists of 8 letters.

```
     1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16
     N   T   O   P   O   S   I   T   I   O   N   S   I   F   T   H   E
(1)  13  22  42  43  42  45  24  22  24  42  13  45  24  15  22  34  33
       D   I   I   H   I   W   O   T   Q   C   H   W   L   W   R   P

     O   N   T   P   O   S   I   T   I   O   N   F   I   F   T   H   R
(2)  42  13  22  43  42  45  24  22  24  42  13  15  24  15  22  34  23
       C   D   I   H   I   W   O   T   Q   C   B   W   L   W   R   O

     W   A   S   P   O   S   I   T   I   O   N   L   I   F   T   I   N
(3)  52  12  45  43  42  45  24  22  24  42  13  41  24  15  22  24  13
       C   I   Y   H   I   W   O   T   Q   C   H   A   L   W   T   L

     T   A   L   P   O   S   I   T   I   O   N   T   I   F   L   I   S
(4)  22  12  41  43  42  45  24  22  24  42  13  22  24  15  41  24  45
       C   I   U   H   I   W   O   T   Q   C   D   T   L   Y   A   Q
       1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16
```

```
1  2  3  4  5  6  7  8  9  10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40    Location
E  W  V  R  V  S  D  D  T  W                                                                                               2I
D  W  V  R  V  S  D  D  T  I                                                                                               2B
R  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  B                                                                             1B
                                       G  U  H  K  I  R  X  Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  W  N           1D
P  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  H  K  I  R  X  Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  Q  D  T  Q     1L
X  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  H  K  I  R  X  Y  I  R  N  B                                                  2D
                                             L  K  I  R  X  Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  Q  D  T  V     1I
                                                            K  I  R  N  G  S  T  L  B  D  D  T  I                          2F
                                                            N  I  R  N  G  S  T  L  B  D  D  T  W                          1G
                                                                              T  W  A  U  U  I  A  A  Q  D                 1H
-----------------------------------------------------------------------------------------------------------------------------
                                                               L  Q  D  T  X  D  W  Y  I  W  X  N  I  N  B  D  R  Q        1H
                                       W  S  A  V  I                                                                       1A
                                       Q  S  A  V  N  I  I  U  H  Q  D  T  X  W  I                                         1E
                                          T  A  V  N  N  I                                                                 1F
                                          Q  A  V  N  I  I  U  H  Q  D  T  X  D  W  Y  I  W  X  N  I  N  B  D  R  W        2G
                                                               L  Q  D  T  X  C  W  S  C  I  V  D  T  N  I  L  V  N        2H
                                                            D  L  Q  D  T  X  W  S  C  L  P                                1K
                                                               C  Q  D  T  Z                                               1J
                                    N  I  L  V  N  O  D  N  U  L  Q  D  U  B  G                                            2I
                              R  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  B  I                                            1B
                              X  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  H  K                                            2E
                                                            L  L  Q  D  U  K                                               2C
-----------------------------------------------------------------------------------------------------------------------------
                                                      T  V  T  I  N  I  N  D  C  U  W  W  D  C                             1C
                                                      W  V  T  I  N  I  N  D  C  U  W  W  D  S  N  I                       1K
-----------------------------------------------------------------------------------------------------------------------------
                                                      D  N  B  A  X                                                        1A
                                                      R  N  B  A  Y  I  I  I                                               2E
                                                      T  E  W  A  Y  I  I  L                                               2E
                                                      I  E  W  A  Y  U  L  Q                                               1E
-----------------------------------------------------------------------------------------------------------------------------
                                       Y  I  I  I  G  T  E  W  A  Y  I                                                     2E
                                                   B  I  E  W  A  Y  U                                                     1E
                                                   Q  V  E  W  C  W  S                                                     1I
                                 I  S  D  C  Y  I  W  V  E  W  D  N  B                                                     1A
                                                B  G  T  E  W  V  R                                                        2I
                                 T  X  D  W  Y  I  W  X  N  I  N                                                           1H
-----------------------------------------------------------------------------------------------------------------------------
                                                A  I  N  B                                                                 1B
                                                K  I  N  D                                                                 2A
                                                W  A  X  L                                                                 2A
                                                B  A  X  W                                                                 2A
```

Figure 81

The pseudo-repetitious letters, I in the third position in cases (1) and (2), and the letters Y_c and U_c in corresponding positions in cases (3) and (4) mean that I, U, and Y, come from the same column of the checkerboard. The I_c in position 2, in cases (1), (3), and (4) and the D_c in the corresponding position in case (2) indicates that I_c and D_c are probably in different columns in the checkerboard. In position 11, the H_c, B_c, and D_c give indications of being in the same row of the checkerboard. In position 14, W_c and Y_c likewise give indications of coming from the same row. But note now that from position 12 it may be deduced that W, A, and T come from the same column of the checkerboard. These are examples of the type of detailed analysis that the student should follow in his attempt to solve a problem of this sort.

l. In general it may be said that when the repetitions are numerous and fairly lengthy, that is when there is a good deal of traffic all in the same checkerboard, and repetitions of tetragraphs and better are plentiful, solution should be relatively easy. In fact, with a fairly large amount of traffic, most of the work involved would consist in listing the 2, 3, 4 . . . letter repetitions. Then a chart would be drawn up to show the *associations* which the prefixes make among themselves and the associations which the suffixes make among themselves. For example:

X A B Q
N A B R
Z A B T
N A B Q
L A B I
Z A B T

Here it is noted that L, N, X, and Z appear as prefixes to repetitions. The letter X is "found in company" with N twice; the "association value" of X and N is 2 units. The association value of Z and N is, however, 4 units, for the N occurs twice and so does the Z. The association value of LX or LZ is 1 unit; that of LN or LZ, 2 units. Thus, the association value for each combination can be studied in all the repetitions and, of course, when the value is high for a given combination it indicates that the two letters really belong together, or in the same column of the checkerboard.

m. What can be done with but one or two relatively short cryptograms depends largely upon their lengths, the number of repetitions they happen to have, the exact construction of the checkerboard, and the ingenuity and patience of the cryptanalyst. Once the letters that constitute the columns and the rows of the checkerboard employed in the recomposition are known, the proper assembling of the columns and rows is a relatively simple matter. If a keyword has been used as the basis for the distribution or mixing of the letters, naturally the reconstruction of the checkerboard is much facilitated. If not, then either the original or an equivalent checkerboard may be reconstructed. Having the recomposition checkerboard at hand, the determination as to whether it is the same as that used in the decomposition follows directly. If not the same, the reconstruction of the decomposition checkerboard is a relatively simple matter.

```
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
  E  W  V  R  V  S  D  D  T  W                                                                                            2I
  D  W  V  R  V  S  D  D  T  I                                                                                            2B
  R  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  B                                                                          1B
                                          G  U  H  K  I  R  X  Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  W  N       1D
  P  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  H  K  I  R  X  Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  Q  D  T  Q  1L
  X  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  H  K  I  R  X  Y  I  R  N  B                                               2D
                                                L  K  I  R  X  Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  Q  D  T  V  1I
                                                            K  I  R  N  G  S  T  L  B  D  D  T  I                         2F
                                                            N  I  R  N  G  S  T  L  B  D  D  T  W                         1G
                                                                              T  W  A  U  U  I  A  A  Q  D                1H
```

Figure 82

47. Example of solution.—*a*. Suppose the two following cryptograms suspected of being in the same key are at hand:

No. 1

Line					
A	ISDCY	IWVEW	DNBAX	WSAVI	WILTK
B	XLYIW	DAINB	DRWVR	VSDDT	NULQD
C	UBINR	KVLQD	NLQDT	VTINI	NDCUW
D	WDCVS	ODCGS	ISSGU	HKIRX	YIRNG
E	TIWAU	UIAAQ	CWNBI	EWAYU	LQSAV
F	NIIUH	QDTXW	ITAVN	NINDC	UWWCG
G	TLGWW	ALWDS	NSILN	NIRNG	STLBD
H	DTWAU	UIAAQ	DGTEL	QDTXD	WYIWX
I	NINBD	RQTQV	EWCWS	CLPIT	LKIRX
J	YIRNG	TIWAU	UIAAQ	CQDTZ	IIWVT
K	ININD	CUWWD	SNILV	NODLQ	DTXWS
L	CLPWV	RVSDD	TNULQ	DUHKI	RXYIR
M	NGTIW	AUUIA	AQCQD	TV	

No. 2

Line					
A	IWILT	GSIHW	WAWKI	NDCUW	WAXLX
B	DIWIR	CVNOD	NGSLN	GIGWL	VFDWV
C	RVSDD	TILLQ	DUKDW	SSHXS	ENCQD
D	TQGTE	UDVQC	OIWTX	WVRVS	DDTNU
E	LQDUH	KIRXY	IRNBA	YIIIG	TEWAY
F	IILWN	KIRNG	STLBD	DTITI	LUDVL
G	VTTAQ	AVNII	UHQDT	XDWYI	WXNIN
H	BDRWS	CLPLW	AHNTL	QDTXC	WSCIV
I	DTNIL	VNODN	ULQDU	BGTEW	VRVSD
J	DTW				

b. A careful and detailed listing of significant repetitions is made, these to show the single-letter prefix and suffix in each case. A partial list of the many repetitions present in the two cryptograms is given in figure 81.

c. Consider the first set of repetitions listed in figure 81, as extracted and shown in figure 82.

PREFIXES

(Letters in same column)

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A										/															
B							𝍸															///			
C				///				//			𝍸 /											/			
D					/		///				/				/		//			///			/		
E															/		/						/		
F																									
G																									
H											𝍸 // 𝍸 ///														
I																			𝍸 /		//				
K													/											////	
L								///												///					
M																									
N																								////	
O																									
P																	/						/		
Q																		//	/			/			
R																							//		
S																			//						
T																						/			
U																									
V																									
W																									
X																									
Y																									
Z																									
	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Figure 83A

According to the principles elucidated in the preceding paragraph, it would seem that the following tentative deductions may be made from the data contained in the columns of figure 82:

(1) From column 1: E, D, R, P, X belong in the same column of the checkerboard.
(2) From column 10: W, I, N belong in the same row.
(3) From column 14: D and G belong in the same column.
(4) From column 16: B and H belong in the same row.
(5) From column 16: H and L belong in the same column.
(6) From column 21: Y, K, N belong in the same column.
(7) From column 25: G and B belong in the same row.
(8) From column 27: I and T belong in the same column.
(9) From column 33: I and W belong in the same row.
(10) From column 36: C and D belong in the same row.
(11) From column 37: W and Q belong in the same row.
(12) From column 40: Q and V belong in the same row.

It would be most fortunate and unusual for all these tentative deductions to be correct, for the disturbing effects of accidental adventitious repetitions have not been taken into account as yet. But let an attempt be made to assemble the data deduced thus far, to see if they can all be reconciled.

d. Tentative deduction (1) indicates that E, D, R, P, and X belong in the same column of the recomposition checkerboard. If correct, the complete set of 5 letters of one column is at hand. But tentative deduction (3) indicates that D and G belong in the same column and this would mean that the column has 6 letters, which is impossible. Further evidence will be required to corroborate the hypothesis that E, D, R, P, and X are all actually in the same column, or that D and G are actually in the same column. For this purpose, further study must be made, and it is convenient to compile an "association table" showing how often certain letters are associated among themselves as prefixes to the repetitions. A similar association table is made for the suffixes. The tables may be combined in a manner similar to that shown in figure 83, where the prefixes to repetitions appear at the left of the central alphabet, the suffixes to the right.

Take column 1 of figure 82, having D, E, P, R, and X as prefixes to a long repetition. A stroke is placed in the E, P, R, and X cells of row D; a stroke is placed in the P, R, and X cells of row E; a stroke is placed in the R and X cells of row P; and finally a stroke is placed in the X cell of row R. Again, take column 16 of figure 82, reading B H H H L. The B need not be considered, since it is not a prefix to the repetition beginning K I R X Y . . ., but the H and L may be considered. In the L cell of row H three strokes are inserted to indicate that H and L are associated that many times. Each time a datum is obtained, it is added to this table. Figure 83 shows the appearance of the table after all the data obtainable from the repetitions listed in subparagraph *b* have been inserted. From even this small amount of material a few deductions can be made. For example, it is seen that the B line of the table for prefixes shows 5 strokes at G and 3 strokes at W, from which it would appear that B, G, and W may be in the same column. The letters C and L likewise seem to be in the same column, as do H and L, making C, H, and L appear to be in the same column. Studying the table of suffixes, it would appear that B and H are in the same row; I and N are in a row. After the entire text has been examined and the prefixes and suffixes distributed in this way, the whole table is studied carefully with a view to eliminating the effects of the accidental or pseudo-repetitious letters, trying to locate those letters which represent the prefixes and suffixes of true repetitions.

SUFFIXES

(Letters in same row)

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A																									
B				/				𝍸		//															
C																		/							
D			/																/						
E																									
F																									
G																									
H										/															
I											/		𝍸/					/		//		/			
K																									
L																						/			
M																									
N																						///			
O																									
P																									
Q																					/	/			
R																									
S																			𝍸/						
T																									
U																									
V																							/		
W																									
X																								///	𝍸
Y																									
Z																									
	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Figure 83B

e. Suppose the data have been reduced to the following:

Letters belonging in same columns		Letters belonging in same rows	
(1)	D, G, U	(1)	G, B, H, K
(2)	H, L, C	(2)	I, W, N, L
(3)	Y, K, N	(3)	D, C, A, S
(4)	W, Q, S, T, B	(4)	Q, V, X, Y, Z
		(5)	T, U

The presumption that Q, V, X, Y, and Z are all in the same row leads to the assumption that the mixing of the checkerboard is based upon a key word or key phrase. Following up this hypothesis, the data are assembled in the following manner:

2d component

1st component	1	2	3	4	5
1	W	I	L	N	
2	T	U			
3	S	D	C	A	
4	B	G	H	K	
5	Q	V	X	Y	Z

FIGURE 84.

f. Only 6 letters remain to be placed in the checkerboard. But there are enough letters already placed to warrant an immediate attempt at decipherment. For example, take the first few groups of message No. 2 and replace the letters by their bipartite equivalents:

I	W	I	L	T	G	S	I	H	W	W	A	W	K	I	N	D
12	11	12	13	21	42	31	12	43	11	11	34	11	44	12	14	32

Recombining the bifid elements:

.1 21 11 21 32 14 23 11 24 31 11 13 41 14 41 21 43 2.

Substituting by means of figure 84:

.1	21	11	21	32	14	23	11	24	31	11	13	41	14	41	21	43	2.
.	T	W	T	D	N	?	W	?	S	W	L	B	N	B	T	H	.

Obviously the decomposition and recomposition checkerboards are different. But the reconstruction of the former is not at all difficult, since the text is now in monoalphabetic form. The message begins with a group showing a repeated letter in the first and third positions: is the 1st word E N E M Y? Probably it is, for message No. 1 also contains the sequence W I L T. At any rate, a transcription of the cryptograms into the bifid equivalents given by the nearly complete recomposition checkerboard (fig. 84) soon yields sufficient monoalphabetic text to permit of the complete reconstruction of both checkerboards:

h. The two cryptograms may now be deciphered directly from the checkerboards. The plain-texts are as follows:

No. 1

```
I  S  D  C  Y  I  W  V  E  W  D  N  B  A  X  W  S  A  V  I  W  I  L  T  K
12 31 32 33 54 12 11 52 24 11 32 14 41 34 53 11 31 34 52 12 11 12 13 21 44
 I  T  I  S  R  E  P  O  R  T  E  D  T  H  A  T  T  H  E  E  N  E  M  Y  H

X  L  Y  I  W  D  A  I  N  B  D  R  W  V  R  V  S  D  D  T  N  U  L  Q  D
53 13 54 12 11 32 34 12 14 41 32 23 11 52 23 52 31 32 32 21 14 22 13 51 32
 A  S  R  E  T  I  R  E  D  T  O  A  P  O  S  I  T  I  O  N  W  E  S  T  O

U  B  I  N  R  K  V  L  Q  D  N  L  Q  D  T  V  T  I  N  I  N  D  C  U  W
22 41 12 14 23 44 52 13 51 32 14 13 51 32 21 52 21 12 14 12 14 32 33 22 11
 F  N  E  W  C  H  E  S  T  E  R  S  T  O  P  O  N  E  R  E  G  I  M  E  N

W  D  C  V  S  O  D  C  G  S  I  S  S  G  U  H  K  I  R  X  Y  I  R  N  G
11 32 33 52 31 15 32 33 42 31 12 31 31 42 22 43 44 12 23 53 54 12 23 14 42
 T  I  S  I  N  V  I  C  I  N  I  T  Y  O  F  C  R  O  S  S  R  O  A  D  O

T  I  W  A  U  U  I  A  A  Q  C  W  N  B  I  E  W  A  Y  U  L  Q  S  A  V
21 12 11 34 22 22 12 34 34 51 33 11 14 41 12 24 11 34 54 22 13 51 31 34 52
 N  E  T  W  O  E  I  G  H  T  A  N  D  N  O  R  T  H  W  E  S  T  T  H  E

N  I  I  U  H  Q  D  T  X  W  I  T  A  V  N  N  I  N  D  C  U  W  W  C  G
14 12 12 22 43 51 32 21 53 11 12 21 34 52 14 14 12 14 32 33 22 11 11 33 42
 R  E  O  F  S  T  O  P  A  N  O  T  H  E  R  R  E  G  I  M  E  N  T  C  O

T  L  G  W  W  A  L  W  D  S  N  S  I  L  N  N  I  R  N  G  S  T  L  B  D
21 13 42 11 11 34 13 11 32 31 14 31 12 13 14 14 12 23 14 42 31 21 13 41 32
 N  C  E  N  T  R  A  T  I  N  G  N  E  A  R  R  O  A  D  J  U  N  C  T  I

D  T  W  A  U  U  I  A  A  Q  D  G  T  E  L  Q  D  T  X  D  W  Y  I  W  X
32 21 11 34 22 22 12 34 34 51 32 42 21 24 13 51 32 21 53 32 11 54 12 11 53
 O  N  T  W  O  E  I  G  H  T  F  O  U  R  S  T  O  P  B  E  P  R  E  P  A

N  I  N  B  D  R  Q  T  Q  V  E  W  C  W  S  C  L  P  I  T  L  K  I  R  X
14 12 14 41 32 23 51 21 51 52 24 11 33 11 31 33 13 45 12 21 13 44 12 23 53
 R  E  D  T  O  S  U  P  P  O  R  T  A  T  T  A  C  K  O  N  C  R  O  S  S

Y  I  R  N  G  T  I  W  A  U  U  I  A  A  Q  C  Q  D  T  Z  I  I  W  V  T
54 12 23 14 42 21 12 11 34 22 22 12 34 34 51 33 51 32 21 55 12 12 11 52 21
 R  O  A  D  O  N  E  T  W  O  E  I  G  H  T  S  T  O  P  K  E  E  P  O  N

I  N  I  N  D  C  U  W  W  D  S  N  I  L  V  N  O  D  L  Q  D  T  X  W  S
12 14 12 14 32 33 22 11 11 32 31 14 12 13 52 14 15 32 13 51 32 21 53 11 31
 E  R  E  G  I  M  E  N  T  I  N  R  E  S  E  R  V  E  S  T  O  P  A  T  T

C  L  P  W  V  R  V  S  D  D  T  N  U  L  Q  D  U  H  K  I  R  X  Y  I  R
33 13 45 11 52 23 52 31 32 32 21 14 22 13 51 32 22 43 44 12 23 53 54 12 23
 A  C  K  P  O  S  I  T  I  O  N  W  E  S  T  O  F  C  R  O  S  S  R  O  A

N  G  T  I  W  A  U  U  I  A  A  Q  C  Q  D  T  V
14 42 21 12 11 34 22 22 12 34 34 51 33 51 32 21 52
 D  O  N  E  T  W  O  E  I  G  H  T  S  T  O  P
```

No. 2

```
I  W  I  L  T  G  S  I  H  W  W  A  W  K  I  N  D  C  U  W  W  A  X  L  X
12 11 12 13 21 42 31 12 43 11 11 34 11 44 12 14 32 33 22 11 11 34 53 13 53
 E  N  E  M  Y  I  N  F  A  N  T  R  Y  R  E  G  I  M  E  N  T  H  A  S  B

D  I  W  I  R  C  V  N  O  D  N  G  S  L  N  G  I  G  W  L  V  F  D  W  V
32 12 11 12 23 33 52 14 15 32 14 42 31 13 14 42 12 42 11 13 52 35 32 11 52
 E  E  N  O  B  S  E  R  V  E  D  I  N  A  D  E  F  E  N  S  I  V  E  P  O

R  V  S  D  D  T  I  L  L  Q  D  U  K  D  W  S  S  H  X  S  E  N  C  Q  D
23 52 31 32 32 21 12 13 13 51 32 22 44 32 11 31 31 43 53 31 24 14 33 51 32
 S  I  T  I  O  N  E  A  S  T  O  F  G  E  T  T  Y  S  B  U  R  G  S  T  O

T  Q  G  T  E  U  D  V  Q  C  O  I  W  T  X  W  V  R  V  S  D  D  T  N  U
21 51 42 21 24 22 32 52 51 33 15 12 11 21 53 11 52 23 52 31 32 32 21 14 22
 P  Y  O  U  W  I  L  L  T  A  K  E  U  P  A  P  O  S  I  T  I  O  N  W  E

L  Q  D  U  H  K  I  R  X  Y  I  R  N  B  A  Y  I  I  I  G  T  E  W  A  Y
13 51 32 22 43 44 12 23 53 54 12 23 14 41 34 54 12 12 12 42 21 24 11 34 54
 S  T  O  F  C  R  O  S  S  R  O  A  D  T  H  R  E  E  F  O  U  R  T  H  R

I  I  L  W  N  K  I  R  N  G  S  T  L  B  D  D  T  I  T  I  L  U  D  V  L
12 12 13 11 14 44 12 23 14 42 31 21 13 41 32 32 21 12 21 12 13 22 32 52 13
 E  E  A  N  D  R  O  A  D  J  U  N  C  T  I  O  N  O  N  E  M  I  L  E  S

V  T  T  A  Q  A  V  N  I  I  U  H  Q  D  T  X  D  W  Y  I  W  X  N  I  N
52 21 21 34 51 34 52 14 12 12 22 43 51 32 21 53 32 11 54 12 11 53 14 12 14
 O  U  T  H  T  H  E  R  E  O  F  S  T  O  P  B  E  P  R  E  P  A  R  E  D

B  D  R  W  S  C  L  P  L  W  A  H  N  T  L  Q  D  T  X  C  W  S  C  I  V
41 32 23 11 31 33 13 45 13 11 34 13 24 11 53 51 32 21 53 33 11 31 33 12 52
 T  O  A  T  T  A  C  K  A  T  D  A  W  N  S  T  O  P  B  A  T  T  A  L  I

D  T  N  I  L  V  N  O  D  N  U  L  Q  D  U  B  G  T  E  W  V  R  V  S  D
32 21 14 12 13 52 14 15 32 14 22 13 51 32 22 41 42 21 24 11 52 23 52 31 32
 O  N  R  E  S  E  R  V  E  W  E  S  T  O  F  Y  O  U  R  P  O  S  I  T  I

D  T  W
32 21 11
 O  N
```

48. Special solution.—*a.* The preceding example of solution constitutes the general solution for this system, since no special conditions are prerequisite to the procedure set forth. An interesting solution, however, is that wherein the same message has been cryptographed by two different sets of checkerboards.

b. Suppose, for instance, that in this system two cryptograms of identical lengths and plain texts but different cryptographic texts are available for examination. They are superimposed and appear as follows:

```
No. 1. G C O D M G C E G B W I L W G M O N G B S X O P C N G E S F L N I
No. 2. W I L T G S I H W W A W K I N D C U W W A X L X D I W I R C V N O

No. 1. W T M G E T L N C G F M D W G X H M G A T A C T O M S W B L G A I
No. 2. D N G S L N G I G W L V F D W V R V S D D T I L L Q D U K D W S S

No. 1. Q P F U A Q M S A Z P H Z G N L M S O W O V X G X H M G A T A K N
No. 2. H X S E N C Q D T Q G T E U D V Q C O I W T X W V R V S D D T N U

No. 1. O M S W C U S H Q L S S T M S U W N N E H U A S U W N T G E L S S
No. 2. L Q D U H K I R X Y I R N B A Y I I I G T E W A Y I I L W N K I R

No. 1. T M O V C V A T A E A C O G N L O P H V S V S U N W T W F M S A X
No. 2. N G S T L B D D T I T I L U D V L V T Q A Q A V N I I U H Q D T X
```

FIRST COMPONENTS OF CRYPTOGRAM No. 1

SECOND COMPONENTS OF CRYPTOGRAM No. 1

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A			T_2I_1 T_2N_1		T_2I_1				S_2S_1	T_2N_1						N_2C_1		W_2A_1 S_2C_1	D_2D_1				T_2X_1 W_2V_1		T_2Q_1
B											K_2D_1							W_2A_1				W_2A_1			
C					I_2H_1		G_2W_1 D_2W_1						D_2I_1	I_2L_1						H_2K_1	L_2B_1	N_2I_1			
D							N_2T_1					T_2G_1	N_2U_1									F_2D_1			
E	I_2T_1						H_2W_1 C_2W_1	G_2T_1			N_2K_1							I_2R_1	L_2N_1						
F											C_2V_1	L_2V_1 H_2Q_1 L_2Q_1								S_2E_1					
G	S_2D_1 W_2S_1	W_2W_1 W_2W_1	W_2I_1 S_2I_1		W_2I_1 S_2L_1 W_2N_1	W_2L_1 T_2L_1						N_2D_1	U_2D_1					W_2A_1		W_2Y_1			W_2V_1 W_2X_1		
H												R_2V_1				R_2X_1				T_2E_1	T_2Q_1				T_2E_1
I																S_2H_1						O_2D_1			
K													N_2U_1		B_2G_1										
L							D_2W_1					V_2Q_1	V_2N_1 G_2I_1	V_2L_1				Y_2I_1 K_2I_1				K_2I_1			
M				V_2F_1			G_2S_1 G_2S_1 V_2S_1							D_2C_1 G_2S_1				Q_2D_1 Q_2C_1 B_2A_1 B_2D_1	V_2D_1						
N			I_2G_1		I_2G_1		U_2W_1 I_2W_1		N_2O_1		D_2V_1	I_2V_1	I_2I_1	U_2L_1					I_2L_1			N_2I_1		L_2P_1	
O				L_2T_1			L_2U_1				L_2V_1	L_2Q_1	C_2U_1		L_2X_1 L_2V_1						W_2T_1 S_2T_1	O_2I_1			
P			X_2D_1	H_2N_1		X_2S_1		G_2T_1 V_2T_1																	
Q											X_2Y_1	C_2Q_1			H_2X_1										
R																									
S	D_2T_1					R_2C_1	R_2W_1 L_2W_1	I_2R_1					C_2L_1 C_2I_1	C_2O_1	A_2H_1			I_2R_1 D_2R_1	R_2N_1	A_2Y_1 A_2V_1	A_2Q_1	D_2U_1	A_2X_1		
T	D_2T_1						L_2W_1				N_2G_1	N_2G_1 N_2B_1		L_2L_1								I_2U_1			
U	E_2N_1 E_2W_1												V_2N_1					K_2I_1				Y_2I_1			
V	B_2D_1		T_2L_1															Q_2A_1					T_2X_1		
W		U_2K_1	U_2H_1	D_2N_1		U_2H_1	I_2N_1 D_2W_1 I_2W_1		A_2W_1	U_2B_1									D_2N_1 I_2I_1 I_2N_1						
X			X_2D_1		X_2C_1		X_2W_1	V_2R_1																	
Y																		P_2L_1							
Z							E_2U_1								Q_2G_1										

Figure 86

No. 1. C G U W G X N W T M S S G A S N Y S G S P D G F M S A X E G A S N
No. 2. D W Y I W X N I N B D R W S C L P L W A H N T L Q D T X C W S C I

No. 1. M T A C W O L N I W D N O M S W K P H U A X H M G A T A C
No. 2. V D T N I L V N O D N U L Q D U B G T E W V R V S D D T I

c. Now consider the first few superimposed letters in these two cryptograms:

No. 1............ G C O D M G C E G . . .
No. 2............ W I L T G S I H W . . .

Take the pair of superimposed letters GW. The G is the cipher resultant of the recombination of two bipartite numerical components that apply to the recomposition checkerboard. The actual identities of these numerical components are not known, but, whatever they be, the first of them determines the first half of G_c, the second determines the second half of G_c. Therefore, for cryptanalytic purposes, the actual, but unknown, numerical components may be represented by the symbols G_1 and G_2, the former referring to the row coordinate of the recomposition checkerboard, the latter to the column coordinate. What has been said of the letter G applies also to the letter W, the equivalent of G in another checkerboard. It will be found that this manner of designating bipartite components by means of subscripts to the letters themselves is a very useful method of handling the letters.

d. Let the first few letters of the two cryptograms be replaced by these same cipher letters with their subscripts to indicate components. Thus:

No. 1............	C	G	C	O	D	M	G	C	E
Components....	C_1C_2	G_1G_2	C_1C_2	O_1O_2	D_1D_2	M_1M_2	G_1G_2	C_1C_2	E_1E_2
No. 2............	I	W	I	L	T	G	S	I	H
Components....	I_1I_2	W_1W_2	I_1I_2	L_1L_2	T_1T_2	G_1G_2	S_1S_2	I_1I_2	H_1H_2

Now from the method of encipherment it is clear that C_2G_1 and I_2W_1 represent the same plain-text letter, since both messages are assumed to contain identical plain texts. That is, C_2G_1 of cryptogram No. 1 $= I_2W_1$ of cryptogram No. 2. Likewise $G_2C_1 = W_2I_2$; $C_2O_1 = I_2L_1$; $O_2D_1 = L_2T_1$; and so on.

e. Let all the component pairs of the cryptograms be equated in this manner and let these pairs be distributed in a table, such as that shown in figure 86. It will be seen in figure 86 that, for example, A_2C_1 of cryptogram No. 1 $= T_2I_1$ and T_2N_1 of cryptogram No. 2. This means that I_1 and N_1 must represent the same *row* coordinate of the recomposition checkerboard for cryptogram No. 2; *in other words* I *and* N *must be in the same row in that checkerboard.* Again, in figure 86, it is seen that $C_2G_1 = G_2W_1$ and D_2W_1, which means that G and D must be in the same column in that checkerboard. Again, $A_2S_1 = W_2A_1 = S_2C_1$; this means that A and C are in the same row, W and S, in the same column, in the recomposition checkerboard for cryptogram No. 2. All these data in figure 86 are studied with the following results:

In same row:

(1) I, N, L, W
(2) A, C, S, D
(3) Q, V, X, Y
(4) B, G

In same column:

(1)	(2)	(3)	(4)	(5)
T	H	Y	Q	U
W	C	K	B	I
S	L			D
	R			G
				V

An attempt is now made to bring together these results to reconstruct the recomposition checkerboard for message No. 2. This yields the following:

2d Component

1st Component

U	K?	R	T	K?
I	N	L	W	K?
D	A	C	S	K?
G	K?	H	B	K?
V	Y?	X	Q	Y?

FIGURE 85*a*.

Compare this with the recomposition checkerboard shown in figure 85 (B). Enough has been shown to illustrate the procedure. If there were just a little more text, probably all 25 letters of the checkerboard could be definitely placed.

f. By making a reciprocal table for equivalencies between component pairs in cryptogram No. 2, the data obtained would permit of reconstructing the recomposition checkerboard for cryptogram No. 1. Having these checkerboards completely or at least partially reconstructed, the reconstruction of the decomposition checkerboards is a relatively easy matter and follows the procedure described in paragraph 47*f*.

g. The complete solution of the two cryptograms, including the decomposition and recomposition matrices, is as follows:

No. 1

2d Component / 1st Component

	1	2	3	4	5
1	R	E	F	L	C
2	T	I	N	G	P
3	O	A	B	D	H
4	K	M	Q	S	U
5	V	W	X	Y	Z

A
(Decomposition)

2d Component / 1st Component

	1	2	3	4	5
1	W	A	S	H	I
2	N	G	T	O	D
3	C	B	E	F	K
4	L	M	P	Q	R
5	U	V	X	Y	Z

B
(Recomposition)

```
 G  C  O  D  M  G  C  E  G  B  W  I  L  W  G  M  O  N  G  B  S  X  O  P  C  N  G  E  S  F
 22 31 24 25 42 22 31 33 22 32 11 15 41 11 22 42 24 21 22 32 13 53 24 43 31 21 22 33 13 34
E  N  E  M  Y  I  N  F  A  N  T  R  Y  R  E  G  I  M  E  N  T  H  A  S  B  E  E  N  O  B

 L  N  I  W  T  M  G  E  T  L  N  C  G  F  M  D  W  G  X  H  M  G  A  T  A  C  T  O  M  S
 41 21 15 11 23 42 22 33 23 41 21 31 22 34 42 25 11 22 53 14 42 22 12 23 12 31 23 24 42 13
S  E  R  V  E  D  I  N  A  D  E  F  E  N  S  I  V  E  P  O  S  I  T  I  O  N  E  A  S  T

 W  B  L  G  A  I  Q  P  F  U  A  Q  M  S  A  Z  P  H  Z  G  N  L  M  S  O  W  O  V  X  G
 11 32 41 22 12 15 44 43 34 51 12 44 42 13 12 55 43 14 55 22 21 41 42 13 24 11 24 52 53 22
O  F  G  E  T  T  Y  S  B  U  R  G  S  T  O  P  Y  O  U  W  I  L  L  T  A  K  E  U  P  A

 X  H  M  G  A  T  A  K  N  O  M  S  W  C  U  S  H  Q  L  S  S  T  M  S  U  W  N  N  E  H
 53 14 42 21 12 23 12 35 21 24 42 13 11 31 51 13 14 44 41 13 13 23 42 13 51 11 21 21 33 14
P  O  S  I  T  I  O  N  W  E  S  T  O  F  C  R  O  S  S  R  O  A  D  T  H  R  E  E  F  O
```

U A S U W N T G E L S S T M O V C V A T A E A C O G N L O P
51 12 13 51 11 21 23 22 33 41 13 13 23 42 24 52 31 52 12 23 12 33 12 31 24 22 21 41 24 43
U R T H R E E A N D R O A D J U N C T I O N O N E M I L E S

H V S V S U N W T W F M S A X C G U W G X N W T M S S G A S
14 52 13 52 13 51 21 11 23 11 34 42 13 12 53 31 22 51 11 22 53 21 11 23 42 13 13 22 12 13
O U T H T H E R E O F S T O P B E P R E P A R E D T O A T T

N Y S G S P D G F M S A X E G A S N M T A C W O L N I W D N
21 54 13 22 13 43 25 22 34 42 13 12 53 33 22 12 13 21 42 23 12 31 11 24 41 21 15 11 25 21
A C K A T D A W N S T O P B A T T A L I O N R E S E R V E W

O M S W K P H U A X H M G A T A C
24 42 13 11 35 43 14 51 12 53 14 42 22 12 23 12 31
E S T O F Y O U R P O S I T I O N

No. 2

		2d Component				
		1	2	3	4	5
1st Component	1	N	U	T	Y	P
	2	E	O	I	F	L
	3	A	M	B	C	S
	4	R	W	G	D	H
	5	K	Q	V	X	Z

		2d Component				
		1	2	3	4	5
1st Component	1	W	I	L	N	O
	2	T	U	R	E	M
	3	S	D	C	A	F
	4	B	G	H	K	P
	5	Q	V	X	Y	Z

W I L T G S I H W W A W K I N D C U W W A X L X D I W I R C
11 12 13 21 42 31 12 43 11 11 34 11 44 12 14 32 33 22 11 11 34 53 13 53 32 12 11 12 23 33
E N E M Y I N F A N T R Y R E G I M E N T H A S B E E N O B

V N O D N G S L N G I G W L V F D W V R V S D D T I L L Q D
52 14 15 32 14 42 31 13 14 42 12 42 11 13 52 35 32 11 52 23 52 31 32 32 21 12 13 13 51 32
S E R V E D I N A D E F E N S I V E P O S I T I O N E A S T

U K D W S S H X S E N C Q D T Q C T E U D V Q C O I W T X W
22 44 32 11 41 31 43 53 31 24 14 33 51 32 21 51 42 21 24 22 32 52 51 33 15 12 11 21 53 11
O F G E T T Y S B U R G S T O P Y O U W I L L T A K E U P A

V R V S D D T N U L Q D U H K I R X Y I R N B A Y I I I G T
52 23 52 31 32 32 21 14 22 13 51 32 22 43 44 12 23 53 54 12 23 14 41 34 54 12 12 12 42 21
P O S I T I O N W E S T O F C R O S S R O A D T H R E E F O

E W A Y I I L W N K I R N G S T L B D D T I T I L U D V L V
24 11 34 54 12 12 13 11 14 44 12 23 14 42 31 21 13 41 32 32 21 12 21 12 13 22 32 52 13 52
U R T H R E E A N D R O A D J U N C T I O N O N E M I L E S

T T A Q A V N I I U H Q D T X D W Y I W X N I N B D R W S C
21 21 34 51 34 52 14 12 12 22 43 51 32 21 53 32 11 54 12 11 53 14 12 14 41 32 23 11 31 33
O U T H T H E R E O F S T O P B E P R E P A R E D T O A T T

L P L W A H N T L Q D T X C W S C I V D T N I L V N O D N U
13 45 13 11 34 13 24 11 53 51 32 21 53 33 11 31 33 12 52 32 21 14 12 13 52 14 15 33 14 22
A C K A T D A W N S T O P B A T T A L I O N R E S E R V E W

L Q D U B G T E W V R V S D D T W
13 51 32 22 41 42 21 24 11 52 23 52 31 32 32 21 11
E S T O F Y O U R P O S I T I O N

h. It is seen that the principles elucidated permit of solving this fairly good cipher system without recourse to frequency studies and detailed, difficult analytical research. What can be done with complete messages of identical texts will give the student a clue to what might be done when fairly lengthy sequences of identical plain texts (but not complete messages) are available for study. Messages with similar beginnings, or similar endings will afford data for such reconstruction.

49. Periodic fractionating systems.—*a.* Another type of combined substitution-transposition system involving fractionation is that in which the processes involved are applied to groupings of fixed length, so that the system gives external evidence of periodicity. One such system, commonly attributed to the French cryptographer Delastelle, is exemplified below. Let the bipartite alphabet be based upon the 25-cell substitution checkerboard shown in figure 80. Let the message to be enciphered be ONE PLANE REPORTED LOST AT SEA. Let it also be assumed that by preagreement between correspondents, periods of 5 letters will constitute the units of encipherment. The bipartite equivalents of the plain-text letters are set down vertically below the letters. Thus:

2d component

1st component	1	2	3	4	5
1	M	A	N	U	F
2	C	T	R	I	G
3	B	D	E	H	K
4	L	O	P	Q	S
5	V	W	X	Y	Z

FIGURE 87.

O N E P L	A N E R E	P O R T E	D L O S T	A T S E A
4 1 3 4 4	1 1 3 2 3	4 4 2 2 3	3 4 4 4 2	1 2 4 3 1
2 3 3 3 1	2 3 3 3 3	3 2 3 2 3	2 1 2 5 2	2 2 5 3 2

Recombinations are effected horizontally within the periods, by joining components in pairs, the first period yielding the pairs 41, 34, 42, 33, 31. These pairs are then replaced by letters from the original checkerboard, yielding the following:

O N E P L	A N E R E	P O R T E	D L O S T	A T S E A
4 1 3 4 4	1 1 3 2 3	4 4 2 2 3	3 4 4 4 2	1 2 4 3 1
2 3 3 3 1	2 3 3 3 3	3 2 3 2 3	2 1 2 5 2	2 2 5 3 2
L H O E B	M D D E E	Q T E R R	H Q T A W	A P A G D

b. A different checkerboard may, of course, be employed for the recomposition process. Also, periods of any convenient length may be employed; or, in a complicated case, periods of varying lengths may be employed in the same cryptogram, according to some prearranged key.

50. General principles underlying the solution.—*a.* It will be noted that the periods in the foregoing example contain an odd number of letters. The result of adopting odd-length periods is to impart a much greater degree of cryptographic security to the system than is the case when even-length periods are involved. This point is worth while elaborating upon to make its cryptanalytic significance perfectly clear. Note what happens when an even period is employed:

O N E P L A	N E R E P O	R T E D L O	. . .
4 1 3 4 4 1	1 3 2 3 4 4	2 2 3 3 4 4	. . .
2 3 3 3 1 2	3 3 3 3 3 2	3 2 3 2 1 2	. . .
L H L R E A	N R Q E E D	T E Q D D A	. . .

Now if each 6-letter cipher group is split in the middle into two sections and the letters are taken alternately from each section (Ex. L H L R E A=L R H E L A) the results are exactly the same as would be obtained in case a simple digraphic encipherment were to be employed with the 2-square checkerboard shown in figure 88. For example, $ON_p=LR_c$; $EP_p=HE_c$, and so on. Encipherment of this sort brings about a fixed relationship between the plain-text digraphs and their cipher equivalents, so that the solution of a message of this type falls under the category of the cryptanalysis of a case of simple digraphic substitution, once the length of the period has been established.[3] The latter step can readily be accomplished, as will be seen presently. In brief, then, it may be said that in this system when encipherment is based upon even periods the cipher text is purely and simply digraphic in character, each plain-text digraph having one and only one cipher-text digraph as its equivalent.

b. But the latter statement is no longer true in the case of odd periods. Note, in the example under paragraph 49*a*, that the cipher equivalent of the first plain-text digraph of the first group, ON, is composed of the initial and final components of the letter L_c, the final component of the letter O_c, and the initial component of the letter L_c. That is, three different plain-text letters, L, O, and E, are involved in the composition of the cipher equivalent of one plain-text digraph, ON. Observe now, in the following examples, that *variants* may be produced for the digraph ON_p.

M	A	N	U	F
C	T	R	I	G
B	D	E	H	K
L	O	P	Q	S
V	W	X	Y	Z
M	C	B	L	V
A	T	D	O	W
N	R	E	P	X
U	I	H	Q	Y
F	G	K	S	Z

O N E P L A N E
L R H E L A N E

FIGURE 88.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
12 34 5	12 34 5	12 34 5	12 34 5	12 34 5	12 34 5	12 34 5
ON EP L	ON TH E	ON CR U	PR ON G	CO NT I	PO NG I	AT IO N
41 34 4	41 23 3	41 22 1	42 41 2	24 12 2	44 12 2	12 24 1
23 33 1	23 24 3	23 13 4	33 23 5	12 32 4	32 35 4	22 42 3
LH OE B	LR DD P	LT AB H	OL RD K	IA CR I	QA RR Y	AI AI R

c. The foregoing examples fall into two classes. In the first, where the O of ON_p falls in an odd position in the period, the first letter of the trigraphic cipher equivalent must be an L_c, the second must be one of the 5 letters in the second column of the substitution checkerboard, the third must be one of the 5 letters in the third row of the checkerboard. Therefore, L_c may combine with 5×5 or 25 pairs of letters to form the second and third letters of the 3-letter equivalent of ON_p. In the other class, where the O of ON_p falls in an even position in the period, the first letter of the equivalent must be one of the 5 letters in the fourth column of the checkerboard, the second must be one of the 5 letters in the first row, and the third letter must be R_c. Therefore, R_c may combine with 5×5 or 25 pairs of letters to form the first and second letters of the 3-letter equivalent of ON_p in this position in the period. Hence, ON_p may be represented by 50 trigraphic combinations; the same is true of all other plain-text digraphs. Now if the system based upon even periods is considered as a simple digraphic substitution, the foregoing remarks lead to characterizing the system based upon odd periods as a special type of digraphic substitution with variants, in which 3 letters represent 2 plain-text letters.

---

[3] An example of the solution of a cryptogram of this type was given in *Military Cryptanalysis, Part I*, sec. IX.

d. However, further study of the odd-period system may show that there is no necessity for trying to handle it as a digraphic system with variants, which would be a rather complex affair. Perhaps the matter can be simplified. Referring again to the example of encipherment in paragraph 49 *a:*

O	N	E	P	L	A	N	E	R	E	P	O	R	T	E	D	L	O	S	T	A	T	S	E	A
4	1	3	4	4	1	1	3	2	3	4	4	2	2	3	3	4	4	4	2	1	2	4	3	1
2	3	3	3	1	2	3	3	3	3	3	2	3	2	3	2	1	2	5	2	2	2	5	3	2
L	H	O	E	B	M	D	D	E	E	Q	T	E	R	R	H	Q	T	A	W	A	P	A	G	D

Now suppose that only the cipher letters are at hand, and that the period is known. The first cipher letter is L, and it is composed of two numerical bifid components that come from the first and second positions in the upper row of components in the period. These components are not known, but whatever they are the first of them is the first component of L, the second of them is the second component of L. Therefore, just as in paragraph 48*c*, the actual but unknown, numerical components may be represented by the symbols L_1 and L_2, the former referring to the row coordinate of the substitution checkerboard, the latter to the column coordinate. The same thing may be done with the components of the second cipher letter, the third, fourth, and fifth, the respective components being placed into their proper positions in the period. Thus:

Cipher	L	H	O	E	B
Components	L_1	L_2	H_1	H_2	O_1
	O_2	E_1	E_2	B_1	B_2

Now let the actual plain-text letters be set into position, as shown at the right in the two diagrams below.

Plain text	O	N	E	P	L
Components	4	1	3	4	4
	2	3	3	3	1
Cipher	L	H	O	E	B

O	N	E	P	L
L_1	L_2	H_1	H_2	O_1
O_2	E_1	E_2	B_1	B_2
L	H	O	E	B

By comparing the two diagrams it becomes obvious that L_1, H_2, and O_1 all represent the coordinate 4; H_1, E_1, E_2, and B_1 all represent the coordinate 3, and so on. If this equivalency were known for all the 50 combinations of the 25 letters with subscript 1 or 2 there would be no problem, for the text of a cryptogram could be reduced to 25 pairs of digits representing monoalphabetic encipherment. But this equivalency is not known in the case of a cryptogram that is to be solved; basically the problem is to establish the equivalency.

e. It is obvious that the vertical pair of components $\genfrac{}{}{0pt}{}{L_1}{O_2}$ represents O_p, the vertical pair $\genfrac{}{}{0pt}{}{L_2}{E_1}$ represents N_p, and so on. The complete example therefore becomes:

Plain	O	N	E	P	L	A	N	E	R	E	P	O	R	T	E	D	L	O	S	T	A	T	S	E	A
Components	L_1	L_2	H_1	H_2	O_1	M_1	M_2	D_1	D_2	D_1	Q_1	Q_2	T_1	T_2	E_1	H_1	H_2	Q_1	Q_2	T_1	A_1	A_2	P_1	P_2	A_1
	O_2	E_1	E_2	B_1	B_2	D_2	E_1	E_2	E_1	E_2	E_2	R_1	R_2	R_1	R_2	T_2	A_1	A_2	W_1	W_2	A_2	G_1	G_2	D_1	D_2
Cipher	L	H	O	E	B	M	D	D	E	E	Q	T	E	R	R	H	Q	T	A	W	A	P	A	G	D

f. Note that a plain-text letter in an odd position in the period has its components in the order $\Theta_1\Theta_2$; in an even position in the period the components of a plain-text letter are in the order $\Theta_2\Theta_1$. For example, note the O_p in the first period $\left(=\genfrac{}{}{0pt}{}{L_1}{O_2}\right)$ and the O_p in the third period $\left(=\genfrac{}{}{0pt}{}{Q_2}{R_1}\right)$. This distinction must be retained since the component indicators for rows and columns are not interchangeable in this system. From this it follows that the vertical pairs of components representing a given plain-text letter are of two classes: $\Theta_1\Theta_2$ and $\Theta_2\Theta_1$, and the two must be kept separate in cryptanalysis.

g. Now consider the equivalent of O_p in the first period. It is composed of $\begin{matrix}L_1\\O_2\end{matrix}$. This is only one of a number of equivalents for O_p in an odd position in the period. The row of the substitution checkerboard indicated by L_1 may be represented by 4 other components, since that row contains 5 letters. Therefore the upper component of the $\begin{matrix}\Theta_1\\\Theta_2\end{matrix}$ equivalent of O_p may be any one of 5 letters. The same is true of the lower component. Hence, O_p in an odd position in the period may be represented by any one of 5×5=25 combinations of vertical components in the sequence $\Theta_1 \rightarrow \Theta_2$. O_p in an even position in the period may be represented by any one of a similar number of combinations of vertical components in the reverse sequence, $\Theta_2 \rightarrow \Theta_1$. Thus, disregarding the position in the period, this system may be described as a monoalphabetic substitution with variants, in which every plain-text letter may be represented by any one of 50 different component-pairs. But in studying an actual cryptogram in this system, since the position (odd or even) occupied by a cipher letter in the period is obvious after the length of the period has been established, a proper segregation of the cipher letters will permit of handling the cipher letters in the two classes referred to above, in which case one has to deal with only 25 variants for each plain-text letter. Obviously, the 25 variants are related to one another by virtue of their having been produced from a single enciphering matrix of but 25 letters. This relationship can be used to good advantage in reconstructing the matrix in the course of the solution and the relationship will be discussed in its proper place.

h. Now if the foregoing encipherment is studied intently several important phenomena may be observed. Note, for instance, how many times either the Θ_1 or the Θ_2 component coincides with the plain-text letter of which it is a part. In the very first period the O_p has an O_2 under it; in the same period the E_p has an E_2 under it. The same phenomenon is observed in columns 3 and 5 of the second period, in column 3 of the third period, and in column 1 of the fifth period. In column 5 of the third, fourth, and fifth periods the Θ_1 components coincide with the respective plain-text letters involved. There are, in this short example, 9 cases of this sort, giving rise to instances of what seems to be a sort of self-encipherment of plain-text letters. How does this come about? And is it an accident that all these cases involve plain-text letters in odd positions in the periods?

i. If the periods in the foregoing example in subparagraph *e* are studied closely, the following observations may be made. Because of the mechanics of encipherment in this system the first cipher letter and the first plain-text letter in each period must come from the same row in the substitution checkerboard. Since there are only 5 letters in a row in the checkerboard the probability that the two letters referred to will be identical is 1/5. (The identity will occur every time that the coordinate of the row in which the second plain-text letter stands in the checkerboard is the same as the coordinate of the column in which the first plain-text letter stands.) The same general remark applies to the second cipher letter and the third plain-text letter; as well as to the third cipher letter and the fifth plain-text letter: In these cases the two letters must come from the same row in the checkerboard and the probability that they will be identical is likewise 1/5. (The identity in the former case will occur every time that the coordinate of the row in which the fourth plain-text letter stands in the checkerboard is the same as that of the column in which the third plain-text letter stands; in the latter case the identity will occur every time that the coordinate of the column in which the first plain-text letter stands is the same as that of the column in which the fifth plain-text letter stands.) The last of the foregoing sources of identity is exemplified in only 4 of the 9 cases mentioned in subparagraph *h* above. These involve the fifth plain-text letter in the third, fourth, and fifth periods, and the first letter in the fifth period, wherein it will be noted that the Θ_1 component standing directly under the plain-text letter is identical with the latter in each case.

j. But how are the other 5 cases of identity brought about? Analysis along the same lines as indicated above will be omitted. It will be sufficient to observe that in each of those cases it is the Θ_2 component which is identical with the plain-text letter involved, and again the probability of the occurrence of the phenomenon in question is 1/5.

k. Since the probability of the occurrence of the event in question is 1/5 for Θ_1 components and 1/5 for Θ_2 components, the total probability from either source of identity is 2/5. This probability applies only to the letters occupying odd positions in the period, and it may be said that in 40 percent of all cases of letters in odd positions in the periods the one or the other of the two cipher components will be identical with the plain-text letter.

l. As regards the plain-text letters in even positions, analysis will show why only in a very few cases will either of the cipher components coincide with the plain-text letter to which they apply, for the method of finding equivalents in the substitution checkerboard is to take the first component as the row coordinate indicator and the second component as the column indicator; a reversal of this order will give wholly different letters, except in those 5 cases in which both components are identical. (The letters involved are those which occupy the 5 cells along the diagonal from the upper left-hand corner to the lower right-hand corner of the checkerboard.) Now in every case of a letter in an odd position in a period the two vertical components are in the $\Theta_1\Theta_2$ order, corresponding to the order in which they are normally taken in finding letter equivalents in the checkerboard. But in every case of a letter in an even position in a period, the two vertical components are in the order $\Theta_2\Theta_1$, which is a reversal of the normal order. It has been seen that in the case of letters in odd positions in the periods the probability that one of the components will coincide with the plain-text letter is 40 percent. The reasoning which led to this determination in the case of the odd letters is exactly the same as that in the case of letters in even positions, except that in the final recomposition process, since the components in the even positions are in the $\Theta_2\Theta_1$ order, which is the reverse of the normal order, identity between one of the components and the plain-text letter can occur in only 1/5 of the 40 = 8 percent of the cases. It may be said then that in this system 48 percent of all the letters of the plain text will be "self-enciphered" and represented by one or the other of the two components; in the case of the letters in odd positions, the amount is 40 percent, in the case of letters in even positions, it is 8 percent.

m. Finally, what of the peculiar phenomenon to be observed in the case of the first column of the fifth period of the example in subparagraph *e*? Here is a case wherein the plain-text value of a pair of superimposed components is unmistakably indicated directly by the cipher components themselves. Studying the cipher group concerned it is noted that it contains 2 A_c's separated by one letter, that is, the A's are 2 intervals apart. This situation is as though the plain-text letter were entirely self-enciphered in this case. Now it is obvious that this phenomenon will occur in the case of periods of 5 letters every time that within a period a cipher letter is repeated at an interval of 2, for this will bring about the superimposition of a Θ_1 and Θ_2 with the same principal letter and therefore the plain-text letter is indicated directly. This question may be pertinent: How many times may this be expected to happen? Analysis along the lines already indicated will soon bring the answer that the phenomenon in question may be expected to happen 4 times out of 100 in the case of letters in odd positions and only 8 times out of 1,000 in the case of letters in even positions. In the latter cases the letters involved are those falling in the diagonal sloping from left to right in the substitution matrix.

n. All of the foregoing phenomena will be useful when the solution of an example is undertaken. But before coming to such an example it is necessary to explain how to ascertain the period of a cryptogram to be solved.

51. Ascertaining the length of the period.—*a.* There are several methods available for ascertaining the length of the period. The simplest, of course, is to look for repetitions of the ordinary sort. If the period is a short one, say 3, 5, 7 letters, and if the message is fairly long, the chances are good that a polygraph which occurs several times within the message will fall in homologous positions within two different periods and therefore will be identically enciphered both times. There will not be many such repetitions, it is true, but factoring the intervals between such as do occur will at least give some clue, if it will not actually disclose the length of the period. For example, suppose that a 7-letter repetition is found, the two occurrences being separated by an interval of 119. The factors of 119 are 7 and 17; the latter is unlikely to be the length of the period, the former, quite likely.

b. If a polygraph is repeated but its two occurrences do not fall in homologous positions in two periods, there will still be manifestations of the presence of repetition but the repeated letters will be separated by one or more intervals in the periods involved. The number of repeated letters will be a function of the length of the polygraph and the length of the period; the interval between the letters constituting the repetition will be a function of the length of the period and the position of the repeated polygraph in two periods in which the two polygraphs occur. Note what happens in the following example:

```
S E N D T H R   E E M E N D O   W N T O E N D   O F E N D I C   O T T R O A D
4 3 1 3 2 3 2   3 3 1 3 1 3 4   5 1 2 4 3 1 3   4 1 3 1 3 2 2   4 2 2 2 4 1 3
5 3 3 2 2 4 3   3 3 1 3 3 2 2   2 3 2 2 3 3 2   2 5 3 3 2 4 1   2 2 2 3 2 2 2
P N R G E T P   E N N P B E T   V I B D D R D   L B D T X D L   O T L D T D T
```

CRYPTOGRAM

```
P N R G E   T P E N N   P B E T V   I B D D R   D L B D T   X D L O T   L D T D T
```

Here the plain text contains the trigraph END 4 times. The END_p in the first period gives rise to the cipher letters $\overset{1}{.}\ \overset{2}{N}\ \overset{3}{.}\ \overset{4}{.}\ \overset{5}{E}\ \overset{6}{.}\ \overset{7}{.}$; in the second period this trigraph also produces $\overset{1}{.}\ \overset{2}{.}\ \overset{3}{N}\ \overset{4}{.}\ \overset{5}{.}\ \overset{6}{E}\ \overset{7}{.}$ The interval between the N_c and the E_c is 3 in both cases. Two times this interval plus one gives the length of the period. In this case the initial letter of the repeated trigraph falls in an even position in the period in both occurrences. The END_p in the third period gives rise to the cipher letters $\overset{1}{.}\ \overset{2}{.}\ \overset{3}{B}\ \overset{4}{.}\ \overset{5}{.}\ \overset{6}{.}\ \overset{7}{D}$; in the fourth period it also produces $\overset{1}{.}\ \overset{2}{B}\ \overset{3}{.}\ \overset{4}{.}\ \overset{5}{.}\ \overset{6}{D}\ \overset{7}{.}$ The interval between the B_c and the D_c is 4 in both cases. Two times this interval minus one gives the length of the period. In this case the initial letter of the repeated trigraph falls in an odd position in the period in both occurrences.

c. The foregoing properties of repetitions in this system afford a means of ascertaining the length of the period in an unknown example. First, it is evident that a repeated trigraph in the plain text produces two different pairs of cipher equivalents according to whether the initial letter of the trigraph occurs in an odd or an even position in the period. The two letters constituting the repetition in the cryptogram will not be sequent but will be separated by an interval of 1, 2, 3, . . . letters depending upon the length of the period. This interval, however, is half of the period plus or minus one.[4] Conversely, if in a cryptogram there are repetitions of pairs of letters separated by an interval x, it is probable that these repetitions represent repetitions of plain-text trigraphs which occupy homologous positions in the period. The interval x (between the letters constituting the repetition in the cipher text) then gives a good clue to the length of the period: p(length of period)$=2x\pm1$.

---

[4] The student must remember that the text is here concerned only with cases in which the period is odd. In the case of even periods the interval separating the 2 letters is always exactly half of the length of the period.

d. A special kind of index is prepared to facilitate the search for repetitions of the nature indicated. If tabulating machinery is available, an alphabetically arranged index showing say 10 succeeding letters after each A_c, B_c, C_c, . . . Z_c is prepared for the cryptogram. Then this index is studied to see how many coincidences occur at various intervals under each letter. For example, under A_c one looks to see if there are 2 or more cases in which the same letter appears 2, 3, 4, . . . intervals to the right of A, a record being kept of the number of such cases under each interval. The same thing is done with reference to B_c, C_c, and so on. The tallies representing coincidences may be amalgamated for all the letters A, B, C, . . . Z, only the intervals being kept segregated. When tabulating machinery is not available, the search for repetitions may be made by transcribing the cryptogram on two long strips of cross-section paper, juxtaposing the strips at A, B, C, . . . Z, and noting the coincidences occurring 1, 2, 3, . . . up to say 10 letters beyond the juxtaposed letters. For example, beginning with A_c, the two strips are juxtaposed with the first A on one against the first A on the other. Note is made of any coincidences found within 10 letters beyond the A's, and a record is kept of such coincidences according to intervals. Keeping one strip in position the other is slid along to the second A, and again coincidences are sought. All the A's are treated in this way, then the B's, C's . . . Z's. The record made of the coincidences may consist merely of a tally stroke written under the intervals 1, 2, 3, . . . 10. That interval which occurs more frequently than all the others is probably the correct one. This interval times 2, plus or minus 1 is the length of the period. There are, therefore, only two alternatives. A choice between the two alternatives may then be made by transcribing the text or a portion of it according to each hypothesis. That transcription which will most often throw the two members constituting a repetition into one and the same period is most likely to be correct.

e. Finally, for ascertaining the period there is one method which is perhaps the most laborious but surest. It has been pointed out that this system reduces to one that may be described as monoalphabetic substitution with variants. If the cipher text is transcribed into Θ_1 and Θ_2 components according to various assumed periods, and then a frequency distribution is made of the pairs of vertical components for each hypothesis, that period which gives the best approximation to the sort of distribution to be expected for a system of monoalphabetic substitution with 25 variants for each letter may be taken to be correct. For in the case of an incorrect period the resultant vertical bipartite components are not the equivalents of the actual plain-text letters; hence such repetitions as occur are purely accidental and the number of such cases would be rather small. But in the case of the correct period the resultant vertical pairs of components are the equivalents of the actual plain-text letters; hence repetitions are causal and fairly frequent. Were it not for variants, of course, the distribution would be perfectly monoalphabetic.

SECOND (θ_2) COMPONENTS

FIRST (θ_1) COMPONENTS

	A	B	C	D	E	F	G	H	I	K	L	M		N	O	P	Q	R	S	T	U	V	W	X	Y	Z		
A		///	/	//			//	////		/			A			/	/			/				/			A	21
B			/	/	/					/	/	/	B		/			//				/					B	15
C			/	///						///	/		C			/	//							/			C	14
D		////			//				/		////		D	/			/	/	/	/				//	/	/	D	23
E	/	/	/				/	//			//		E	/		/		/				/					E	14
F									//		/		F	/						/							F	6
G						//							G		///			/					//				G	8
H			/	/					/		//		H											/			H	8
I	/	/			/						//		I	/	/												I	8
K		///	/	/	/	//							K			/		/	/					/		/	K	13
L	//								/	/	/		L		///		//	///	/			/			/		L	17
M		/			//	/	//	/		/	////		M	/		/	/						/	/			M	16
	A	B	C	D	E	F	G	H	I	K	L	M		N	O	P	Q	R	S	T	U	V	W	X	Y	Z		
N	/			/					/	////	///		N		/		/	/				/					N	14
O												/	O			/		//	/				/		/		O	7
P	/		/							/		/	P					/					//				P	7
Q		//	/			/							Q		//								//				Q	8
R										//		/	R					//	/					/		/	R	8
S		///	/										S		/	//		/	/								S	9
T	//				/		/	/					T							////				/			T	11
U	//		//	/	////		/	//					U									/	/	/			U	15
V	/				/			/	/		///		V							/							V	8
W								/					W					//		/				///			W	7
X		/	//	/						/	/		X		///	/		////	/				/				X	16
Y			//	/						//		////	Y		//		/						//				Y	14
Z					////					/			Z		/					//				/			Z	9
	A 12	B 19	C 15	D 14	E 17	F 7	G 9	H 12	I 7	K 18	L 26	M 8		N 5	O 18	P 9	Q 9	R 28	S 7	T 13	U 0	V 5	W 12	X 17	Y 9	Z 4	(Total=296)	296

Figure 90

FIRST (θ_1) COMPONENTS

SECOND (θ_2) COMPONENTS

	A	B	C	D	E	F	G	H	I	K	L	M		N	O	P	Q	R	S	T	U	V	W	X	Y	Z		
A	/	/	//	///		/				//	//	//	A		/					/		/	/	//			A	20
B		/							/		///		B		//	/								/	/		B	10
C							/				//		C		/		/	/		///					/		C	10
D	///	//		//		//	/	///		//		/	D	//													D	18
E	/	/											E				/	///		/		/	/				E	9
F					//						/		F											/	/		F	5
G					/			/		/			G				//	/									G	6
H										//			H				/		/								H	4
I	/				/						/		I		/											/	I	5
K	/			/	//		/						K		/									/			K	7
L		/			//								L			///	/	//		/			/	/			L	12
M								/		/			M				//	///		///			/			/	M	12
	A	B	C	D	E	F	G	H	I	K	L	M		N	O	P	Q	R	S	T	U	V	W	X	Y	Z		
N		//	//	/	/	/	/	/	/		/		N					/						/			N	13
O							//						O									/					O	3
P			/					/					P		/			/									P	4
Q					/	/					///		Q					/							/		Q	7
R													R					/		//			//				R	5
S											/		S		///			/						/	/		S	7
T		/	/							//			T					//	///								T	9
U	/		/	//	/	/			/		//	/	U		//	//									/		U	15
V				/	/								V	/										//	/		V	6
W										/			W					/					/				W	3
X	/									//			X	/				/	//				//			/	X	10
Y		𝍸	///					/			/		Y	/	/			/									Y	13
Z			/				/		/		////		Z		/												Z	8
	A	B	C	D	E	F	G	H	I	K	L	M		N	O	P	Q	R	S	T	U	V	W	X	Y	Z		221
	9	14	11	10	12	6	7	8	4	13	21	4		5	14	6	8	20	6	11	0	3	9	10	7	3	(Total=221)	

Figure 91

52. Illustration of solution.—*a.* With the foregoing principles in mind, the following cryptogram will be studied:

```
KZFBE ILYYM OCBRB LZDOT GBLPK YWCUC CEPQL
AMEYL ZQXWH LRWQY DRWBM TIZEB ELAYE SOBRY
QVBBL YXNAB QBDOY MQDLW LNACO XCRRG ASWQB
FDDTE BAMFD ETENA KGDFO QDUBN DCLYD VWBAX
CAUGG XOART XXTSD AYXHK OLSXA BRKRP UZWHO
MTDHT SGMLS LQPOU NHCIC KKAQB DOFLE KAPRG
SXUPO WALMA VQHLM LAXKP WSTMC XKQVH SIXSL
LWXLX RSGZD FKLNY BXMRB NADKT TBAEO BHWVL
YSXMB OWPGX KORZI UCEAD YIDBL ZMITA NHCAI
DNCID DOYIB CNOLY UUMCE POTDM GBFUN AHLBD
WXNXK KCSCT OXTSD AYXHK CNLDK RRFAY APMHC
ANMBV GREZQ ATCYI MNDLR LGMTW ETRCV VKTED
UFDEL XHEQV CBLYU DUGYA FHNQL KFRUC NVDLH
LZDRE LKXKU PSEMC TNKTK EBOEE PGVQT GWERH
LZDRE LKFAX IYDAK ZLXXO RRPER RRRNC IE
```

b. The long repetitions noted in the text (intervals=210 and 35) indicate a period of either 5 or 7. By transcribing several lines of text into their Θ_1 and Θ_2 components according to both of these alternatives and distributing the vertically superimposed pairs, it is soon found that a period of 7 produces many more repetitions than does a period of 5. The entire text is then transcribed into its Θ_1 and Θ_2 components according to a period of 7 (see fig. 89) and complete distributions of $\Theta_1\Theta_2$ and $\Theta_2\Theta_1$ vertical pairs are made, the distributions being, of course, kept separate. They are shown in figures 90 and 91. The individual distributions show many repetitions and the distributions as a whole are very favorable for a period of 7.

1	2	3	4	5
K Z F B E I L	Y Y M O C B R	B L Z D O T G	B L P K Y W C	U C C E P Q L
$K_1 K_2 Z_1 Z_2 F_1 F_2 B_1$	$Y_1 Y_2 Y_1 Y_2 M_1 M_2 O_1$	$B_1 B_2 L_1 L_2 Z_1 Z_2 D_1$	$B_1 B_2 L_1 L_2 P_1 P_2 K_1$	$U_1 U_2 C_1 C_2 C_1 C_2 E_1$
$B_2 E_1 E_2 I_1 I_2 L_1 L_2$	$O_2 C_1 C_2 B_1 B_2 R_1 R_2$	$D_2 O_1 O_2 T_1 T_2 G_1 G_2$	$K_2 Y_1 Y_2 W_1 W_2 C_1 C_2$	$E_2 P_1 P_2 Q_1 Q_2 L_1 L_2$
6	7	8	9	10
A M E Y L Z Q	X W H L R W Q	Y D R W B M T	I Z E B E L A	Y E S O B R Y
$A_1 A_2 M_1 M_2 E_1 E_2 Y_1$	$X_1 X_2 W_1 W_2 H_1 H_2 L_1$	$Y_1 Y_2 D_1 D_2 R_1 R_2 W_1$	$I_1 I_2 Z_1 Z_2 E_1 E_2 B_1$	$Y_1 Y_2 E_1 E_2 S_1 S_2 O_1$
$Y_2 L_1 L_2 Z_1 Z_2 Q_1 Q_2$	$L_2 R_1 R_2 W_1 W_2 Q_1 Q_2$	$W_2 B_1 B_2 M_1 M_2 T_1 T_2$	$B_2 E_1 E_2 L_1 L_2 A_1 A_2$	$O_2 B_1 B_2 R_1 R_2 Y_1 Y_2$
11	12	13	14	15
Q V B B L Y X	N A B Q B D O	Y M Q D L W L	N A C O X C R	R G A S W Q B
$Q_1 Q_2 V_1 V_2 B_1 B_2 B_1$	$N_1 N_2 A_1 A_2 B_1 B_2 Q_1$	$Y_1 Y_2 M_1 M_2 Q_1 Q_2 D_1$	$N_1 N_2 A_1 A_2 C_1 C_2 O_1$	$R_1 R_2 G_1 G_2 A_1 A_2 S_1$
$B_2 L_1 L_2 Y_1 Y_2 X_1 X_2$	$Q_2 B_1 B_2 D_1 D_2 O_1 O_2$	$D_2 L_1 L_2 W_1 W_2 L_1 L_2$	$O_2 X_1 X_2 C_1 C_2 R_1 R_2$	$S_2 W_1 W_2 Q_1 Q_2 B_1 B_2$
16	17	18	19	20
F D D T E B A	M F D E T E N	A K G D F O Q	D U B N D C L	Y D V W B A X
$F_1 F_2 D_1 D_2 D_1 D_2 T_1$	$M_1 M_2 F_1 F_2 D_1 D_2 E_1$	$A_1 A_2 K_1 K_2 G_1 G_2 D_1$	$D_1 D_2 U_1 U_2 B_1 B_2 N_1$	$Y_1 Y_2 D_1 D_2 V_1 V_2 W_1$
$T_2 E_1 E_2 B_1 B_2 A_1 A_2$	$E_2 T_1 T_2 E_1 E_2 N_1 N_2$	$D_2 F_1 F_2 O_1 O_2 Q_1 Q_2$	$N_2 D_1 D_2 C_1 C_2 L_1 L_2$	$W_2 B_1 B_2 A_1 A_2 X_1 X_2$

FIGURE 89.

21
CAUGGXO
$C_1 C_2 A_1 A_2 U_1 U_2 G_1$
$G_2 G_1 G_2 X_1 X_2 O_1 O_2$

22
ARTXXTS
$A_1 A_2 R_1 R_2 T_1 T_2 X_1$
$X_2 X_1 X_2 T_1 T_2 S_1 S_2$

23
DAYXHKO
$D_1 D_2 A_1 A_2 Y_1 Y_2 X_1$
$X_2 H_1 H_2 K_1 K_2 O_1 O_2$

24
LSXABRK
$L_1 L_2 S_1 S_2 X_1 X_2 A_1$
$A_2 B_1 B_2 R_1 R_2 K_1 K_2$

25
RPUZWHO
$R_1 R_2 P_1 P_2 U_1 U_2 Z_1$
$Z_2 W_1 W_2 H_1 H_2 O_1 O_2$

26
MTDHTSG
$M_1 M_2 T_1 T_2 D_1 D_2 H_1$
$H_2 T_1 T_2 S_1 S_2 G_1 G_2$

27
MLSLQPO
$M_1 M_2 L_1 L_2 S_1 S_2 L_1$
$L_2 Q_1 Q_2 P_1 P_2 O_1 O_2$

28
UNHCICK
$U_1 U_2 N_1 N_2 H_1 H_2 C_1$
$C_2 I_1 I_2 C_1 C_2 K_1 K_2$

29
KAQBDOF
$K_1 K_2 A_1 A_2 Q_1 Q_2 B_1$
$B_2 D_1 D_2 O_1 O_2 F_1 F_2$

30
LEKAPRG
$L_1 L_2 E_1 E_2 K_1 K_2 A_1$
$A_2 P_1 P_2 R_1 R_2 G_1 G_2$

31
SXUPOWA
$S_1 S_2 X_1 X_2 U_1 U_2 P_1$
$P_2 O_1 O_2 W_1 W_2 A_1 A_2$

32
LMAVQHL
$L_1 L_2 M_1 M_2 A_1 A_2 V_1$
$V_2 Q_1 Q_2 H_1 H_2 L_1 L_2$

33
MLAXKPW
$M_1 M_2 L_1 L_2 A_1 A_2 X_1$
$X_2 K_1 K_2 P_1 P_2 W_1 W_2$

34
STMCXKQ
$S_1 S_2 T_1 T_2 M_1 M_2 C_1$
$C_2 X_1 X_2 K_1 K_2 Q_1 Q_2$

35
VHSIXSL
$V_1 V_2 H_1 H_2 S_1 S_2 I_1$
$I_2 X_1 X_2 S_1 S_2 L_1 L_2$

36
LWXLXRS
$L_1 L_2 W_1 W_2 X_1 X_2 L_1$
$L_2 X_1 X_2 R_1 R_2 S_1 S_2$

37
GZDFKLN
$G_1 G_2 Z_1 Z_2 D_1 D_2 F_1$
$F_2 K_1 K_2 L_1 L_2 N_1 N_2$

38
YBXMRBN
$Y_1 Y_2 B_1 B_2 X_1 X_2 M_1$
$M_2 R_1 R_2 B_1 B_2 N_1 N_2$

39
ADKTTBA
$A_1 A_2 D_1 D_2 K_1 K_2 T_1$
$T_2 T_1 T_2 B_1 B_2 A_1 A_2$

40
EOBHWVL
$E_1 E_2 O_1 O_2 B_1 B_2 H_1$
$H_2 W_1 W_2 V_1 V_2 L_1 L_2$

41
YSXMBOW
$Y_1 Y_2 S_1 S_2 X_1 X_2 M_1$
$M_2 B_1 B_2 O_1 O_2 W_1 W_2$

42
PGXKORZ
$P_1 P_2 G_1 G_2 X_1 X_2 K_1$
$K_2 O_1 O_2 R_1 R_2 Z_1 Z_2$

43
IUCEADY
$I_1 I_2 U_1 U_2 C_1 C_2 E_1$
$E_2 A_1 A_2 D_1 D_2 Y_1 Y_2$

44
IDBLZMI
$I_1 I_2 D_1 D_2 B_1 B_2 L_1$
$L_2 Z_1 Z_2 M_1 M_2 I_1 I_2$

45
TANHCAI
$T_1 T_2 A_1 A_2 N_1 N_2 H_1$
$H_2 C_1 C_2 A_1 A_2 I_1 I_2$

46
DNCIDDO
$D_1 D_2 N_1 N_2 C_1 C_2 I_1$
$I_2 D_1 D_2 D_1 D_2 O_1 O_2$

47
YIBCNOL
$Y_1 Y_2 I_1 I_2 B_1 B_2 C_1$
$C_2 N_1 N_2 O_1 O_2 L_1 L_2$

48
YUUMCEP
$Y_1 Y_2 U_1 U_2 U_1 U_2 M_1$
$M_2 C_1 C_2 E_1 E_2 P_1 P_2$

49
OTDMGBF
$O_1 O_2 T_1 T_2 D_1 D_2 M_1$
$M_2 G_1 G_2 B_1 B_2 F_1 F_2$

50
UNAHLBD
$U_1 U_2 N_1 N_2 A_1 A_2 H_1$
$H_2 L_1 L_2 B_1 B_2 D_1 D_2$

51
WXNXKKC
$W_1 W_2 X_1 X_2 N_1 N_2 X_1$
$X_2 K_1 K_2 K_1 K_2 C_1 C_2$

52
SCTOXTS
$S_1 S_2 C_1 C_2 T_1 T_2 O_1$
$O_2 X_1 X_2 T_1 T_2 S_1 S_2$

53
DAYXHKC
$D_1 D_2 A_1 A_2 Y_1 Y_2 X_1$
$X_2 H_1 H_2 K_1 K_2 C_1 C_2$

54
NLDKRRF
$N_1 N_2 L_1 L_2 D_1 D_2 K_1$
$K_2 R_1 R_2 R_1 R_2 F_1 F_2$

55
KYAPMHC
$K_1 K_2 Y_1 Y_2 A_1 A_2 P_1$
$P_2 M_1 M_2 H_1 H_2 C_1 C_2$

56
ANMBVGR
$A_1 A_2 N_1 N_2 M_1 M_2 B_1$
$B_2 V_1 V_2 G_1 G_2 R_1 R_2$

57
EZQATCY
$E_1 E_2 Z_1 Z_2 Q_1 Q_2 A_1$
$A_2 T_1 T_2 C_1 C_2 Y_1 Y_2$

58
IMNDLRL
$I_1 I_2 M_1 M_2 N_1 N_2 D_1$
$D_2 L_1 L_2 R_1 R_2 L_1 L_2$

59
GMTWETR
$G_1 G_2 M_1 M_2 T_1 T_2 W_1$
$W_2 E_1 E_2 T_1 T_2 R_1 R_2$

60
CVVKTED
$C_1 C_2 V_1 V_2 V_1 V_2 K_1$
$K_2 T_1 T_2 E_1 E_2 D_1 D_2$

61
UFDELXH
$U_1 U_2 F_1 F_2 D_1 D_2 E_1$
$E_2 L_1 L_2 X_1 X_2 H_1 H_2$

62
EQVCBLY
$E_1 E_2 Q_1 Q_2 V_1 V_2 C_1$
$C_2 B_1 B_2 L_1 L_2 Y_1 Y_2$

63
UDUGYAF
$U_1 U_2 D_1 D_2 U_1 U_2 G_1$
$G_2 Y_1 Y_2 A_1 A_2 F_1 F_2$

64
HNQLKFR
$H_1 H_2 N_1 N_2 Q_1 Q_2 L_1$
$L_2 K_1 K_2 F_1 F_2 R_1 R_2$

65
UCNVDLH
$U_1 U_2 C_1 C_2 N_1 N_2 V_1$
$V_2 D_1 D_2 L_1 L_2 H_1 H_2$

66
LZDRELK
$L_1 L_2 Z_1 Z_2 D_1 D_2 R_1$
$R_2 E_1 E_2 L_1 L_2 K_1 K_2$

67
XKUPSEM
$X_1 X_2 K_1 K_2 U_1 U_2 P_1$
$P_2 S_1 S_2 E_1 E_2 M_1 M_2$

68
CTNKTKE
$C_1 C_2 T_1 T_2 N_1 N_2 K_1$
$K_2 T_1 T_2 K_1 K_2 E_1 E_2$

69
BOEEPGV
$B_1 B_2 O_1 O_2 E_1 E_2 E_1$
$E_2 P_1 P_2 G_1 G_2 V_1 V_2$

70
QTGWERH
$Q_1 Q_2 T_1 T_2 G_1 G_2 W_1$
$W_2 E_1 E_2 R_1 R_2 H_1 H_2$

71
LZDRELK
$L_1 L_2 Z_1 Z_2 D_1 D_2 R_1$
$R_2 E_1 E_2 L_1 L_2 K_1 K_2$

72
FAXIYDA
$F_1 F_2 A_1 A_2 X_1 X_2 I_1$
$I_2 Y_1 Y_2 D_1 D_2 A_1 A_2$

73
KZLXXOR
$K_1 K_2 Z_1 Z_2 L_1 L_2 X_1$
$X_2 X_1 X_2 O_1 O_2 R_1 R_2$

74
RPERRRR
$R_1 R_2 P_1 P_2 E_1 E_2 R_1$
$R_2 R_1 R_2 R_1 R_2 R_1 R_2$

75
NCIE
$N_1 N_2 C_1 C_2$
$I_1 I_2 E_1 E_2$

FIGURE 89—Continued.

c. The text now being transcribed into periods of 7, with the Θ_1 and Θ_2 components indicated by the cipher letters in each period, the vertical pairs of components are examined to locate cases in which the basic letters of the Θ_1 and Θ_2 superimposed components are identical, whereupon the plain-text letters indicated are at once inserted into position. In this example 10 such cases are found, one each in periods 14, 22, 26, 35, 36, 52, 59, 68, and two in period 74. All of these, of course, involve letters in odd positions in the periods. The plain-text letters thus inserted may serve as clues for assuming probable words.

d. Now if only a few equivalencies can be established between a few of the Θ_1 components, or between a few of the Θ_2 components, or between a few Θ_1 and Θ_2 components, a long step forward may be taken in the solution. Perhaps some information can be found by studying figures 90 and 91. A consideration of figure 90 will soon lead to the idea that each row of frequencies can indicate only 5 different plain-text letters, one of which coincides with the indicating letter at the left of the row. Moreover, in this same figure, while there are 25 rows in all, there are really only 5 different *categories* of rows, each category corresponding to a row in the substitution checkerboard.

e. To explain quite clearly what is meant and how the principle can be employed in this case, assume that figure 90, instead of applying to an unknown checkerboard, applied to a known one, say that shown in figure 87. The bipartite coordinates and the letters which would occupy the cells are as seen in figure 92:

		2	1	1	2	3	5	5	4	4	5	1	1	3	2	3	4	3	5	2	4	1	2	3	4	5
		A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	A	A	M	M	A	N	F	F	U	U	F	M	M	N	A	N	U	N	F	A	U	M	A	N	U	F
3	B	D	B	B	D	E	K	K	H	H	K	B	B	E	D	E	H	E	K	D	H	B	D	E	H	K
2	C	T	C	C	T	R	G	G	I	I	G	C	C	R	T	R	I	R	G	T	I	C	T	R	I	G
3	D	D	B	B	D	E	K	K	H	H	K	B	B	E	D	E	H	E	K	D	H	B	D	E	H	K
3	E	D	B	B	D	E	K	K	H	H	K	B	B	E	D	E	H	E	K	D	H	B	D	E	H	K
1	F	A	M	M	A	N	F	F	U	U	F	M	M	N	A	N	U	N	F	A	U	M	A	N	U	F
2	G	T	C	C	T	R	G	G	I	I	G	C	C	R	T	R	I	R	G	T	I	C	T	R	I	G
3	H	D	B	B	D	E	K	K	H	H	K	B	B	E	D	E	H	E	K	D	H	B	D	E	H	K
2	I	T	C	C	T	R	G	G	I	I	G	C	C	R	T	R	I	R	G	T	I	C	T	R	I	G
3	K	D	B	B	D	E	K	K	H	H	K	B	B	E	D	E	H	E	K	D	H	B	D	E	H	K
4	L	O	L	L	O	P	S	S	Q	Q	S	L	L	P	O	P	Q	P	S	O	Q	L	O	P	Q	S
1	M	A	M	M	A	N	F	F	U	U	F	M	M	N	A	N	U	N	F	A	U	M	A	N	U	F
1	N	A	M	M	A	N	F	F	U	U	F	M	M	N	A	N	U	N	F	A	U	M	A	N	U	F
4	O	O	L	L	O	P	S	S	Q	Q	S	L	L	P	O	P	Q	P	S	O	Q	L	O	P	Q	S
4	P	O	L	L	O	P	S	S	Q	Q	S	L	L	P	O	P	Q	P	S	O	Q	L	O	P	Q	S
4	Q	O	L	L	O	P	S	S	Q	Q	S	L	L	P	O	P	Q	P	S	O	Q	L	O	P	Q	S
2	R	T	C	C	T	R	G	G	I	I	G	C	C	R	T	R	I	R	G	T	I	C	T	R	I	G
4	S	O	L	L	O	P	S	S	Q	Q	S	L	L	P	O	P	Q	P	S	O	Q	L	O	P	Q	S
2	T	T	C	C	T	R	G	G	I	I	G	C	C	R	T	R	I	R	G	T	I	C	T	R	I	G
1	U	A	M	M	A	N	F	F	U	U	F	M	M	N	A	N	U	N	F	A	U	M	A	N	U	F
5	V	W	V	V	W	X	Z	Z	Y	Y	Z	V	V	X	W	X	Y	X	Z	W	Y	V	W	X	Y	Z
5	W	W	V	V	W	X	Z	Z	Y	Y	Z	V	V	X	W	X	Y	X	Z	W	Y	V	W	X	Y	Z
5	X	W	V	V	W	X	Z	Z	Y	Y	Z	V	V	X	W	X	Y	X	Z	W	Y	V	W	X	Y	Z
5	Y	W	V	V	W	X	Z	Z	Y	Y	Z	V	V	X	W	X	Y	X	Z	W	Y	V	W	X	Y	Z
5	Z	W	V	V	W	X	Z	Z	Y	Y	Z	V	V	X	W	X	Y	X	Z	W	Y	V	W	X	Y	Z

FIGURE 92.

Now consider the **A** row and the **F** row. The 25 letters in both rows of cells are, from the very nature of the system, identical in their sequence and there are only 5 different letters involved, each appearing 5 times. Therefore it would seem that frequency distributions corresponding to these rows should show definite characteristics by means of which they could be compared statistically. Furthermore, the Θ_1 coordinates applying to these two rows, A_1 and F_1, indicate that **A** and **F** are in the same row in the checkerboard. What has been said of the **A** and **F** rows also applies to the **M**, **N**, and **U** rows, for the letters **A**, **F**, **M**, **N**, and **U** are all in the same row in the checkerboard (fig. 87). Perhaps a statistical test can be applied to ascertain which rows of distributions in figure 90 are similar and this in turn may give clues to the letters which fall in the same row in the checkerboard applicable to the problem in hand.

f. Again, consider the columns in figure 90. What has been said of the rows applies equally to the columns, and therefore the same sort of test may also be applied to the columns of figure 90 for clues as to the composition of the columns of the checkerboard applicable to the problem under consideration. If there were sufficient text much of the labor of solving such cases would be reduced to a matter of statistical analysis. But what sort of statistical test should be used? Obviously it should be one based upon "matching" the distributions of figure 90, but specifically what should it be? Note the distributions in rows **D** and **M**; they appear to be similar. Is it correct to apply the usual χ-test for matching two frequency distributions? Consider the composition of the rows of figure 90, and specifically consider the **A** and **F** rows, composed as follows:

A.................. A M M A N F F U U F M M N A N U N F A U M A N U F
F.................. A M M A N F F U U F M M N A N U N F A U M A N U F

Here the letters in opposite cells are identical and there are only 5 different letters involved: **A**, **M**, **N**, **F**, and **U**. Of these only 3 are high-frequency letters in normal plain text; 2 are of medium to low frequency. But the high-frequency letters in the **A** row match those in the **F** row, the low-frequency letters in the **A** row also match the low-frequency letters in the **F** row. Hence, if frequency distributions corresponding to these rows are tested statistically, they should yield a fairly high index of coincidence. But should the constant .0667 (probability of monographic coincidence in normal English text) be used in the test? Obviously not, for this constant is derived from statistics based upon the normal frequencies of all 26 letters of the alphabet, whereas here only 5 letters are involved and the exact 5 involved in any example is determined by the composition of the checkerboard. Again, consider the **A** and **C** rows of figure 90, composed as follows:

A................ A M M A N F F U U F M M N A N U N F A U M A N U F
C................ T C C T R G G I I G C C R T R I R G T I C T R I G

Here is a case where, by chance, high-frequency letters stand opposite high-frequency letters (**A** and **T**, **N** and **R**); medium-frequency letters stand opposite medium-frequency letters (**M** and **C**, **F** and **G**). The only case of fairly marked difference is in that of the pairing of **U** and **I**. Hence, a statistical matching of frequency distributions applying to these two rows would be apt to yield a high index of coincidence. Yet, these two rows do not belong together and to assume that the letters **A** and **C** belong in the same row in the checkerboard would block or at least retard solution. In spite of the foregoing reasoning, there nevertheless remains the feeling that a statistical matching of the rows should be possible or should at least offer some clues as to the composition of the checkerboard.

g. In applying the usual χ-test for matching two distributions use is made of the important constant .0667, the probability of monographic coincidence for normal English text. This constant may be modified to meet the special conditions of the present problem. If it be assumed that the mixing of the letters in the checkerboard is fairly good, in normal cases it may be assumed that there will be 1 high-frequency letter, 3 medium-frequency letters, and 1 low-frequency letter in each row and in each column of the checkerboard. Suppose the letters in each category be as follows:

High frequency	A E I N O R S T
Medium frequency	B C D F G H L M P U Y
Low frequency	K Q V W X Z

Adding the squares of the probabilities for separate occurrence [5] of the letters in each category:

A	.0054
E	.0169
I	.0054
N	.0063
O	.0057
R	.0057
S	.0037
T	.0084
Total=	.0575
Average=	.0072

B	.0001
C	.0009
D	.0018
F	.0008
G	.0003
H	.0012
L	.0013
M	.0006
P	.0007
U	.0007
Y	.0004
Total=	.0088
Average=	.0008

K	.0000
Q	.0000
V	.0002
W	.0002
X	.0000
Z	.0000
Total=	.0004
Average=	.00007

Since each row of figure 90 contains 25 letters, composed of 5 different letters each appearing 5 times, and it is assumed that each row of the checkerboard contains 2 high-frequency letters, 2 medium-frequency letters, and 1 low-frequency letter, the rows in figure 90 will be composed of 10 high-frequency letters, 10 medium-frequency letters, and 5 low-frequency letters. Therefore, the sum of the squares of the average probabilities of the letters occurring in each row of figure 90 is as follows:

$$5 \times .0072 = .0360$$
$$15 \times .0008 = .0120$$
$$5 \times .00007 = .0004$$
$$\text{Total} = .0484$$

This, then, is the constant that should be applied in the χ-test for the problem under consideration. Suppose, for convenience, the approximation .05 is used. This is considerably less than the normal constant .0667 and means that in the case of this problem two distributions can be considered to "match" even if the number of coincidences (value of χ) is considerably less than what would be expected in the case of the normal type of frequency distribution. However, it must be remembered that even if two distributions give an observed value for χ that is close to or even greater than the expected, one can still not be certain that the two distributions apply to identical rows of letters and indicate two letters in the same row in the checkerboard, since it may happen that the composition of the checkerboard is such that two rows have letters of about the same frequency values, as pointed out above.

h. With this reservation in mind, let figure 90 be examined. Take rows D and M, which on casual examination look a good deal alike, as seen in figure 93.

---

[5] As given in the table on p. 114 of *Military Cryptanalysis*, Part 1, Appendix 2, par. 2*e* (1).

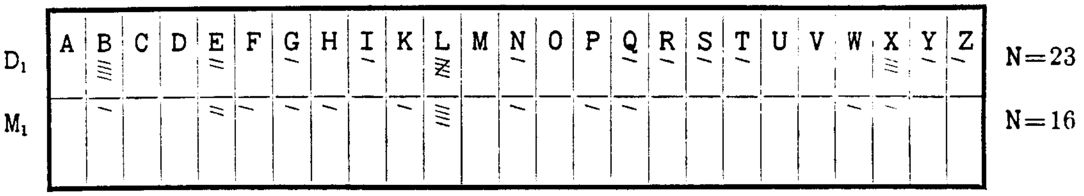


FIGURE 93.

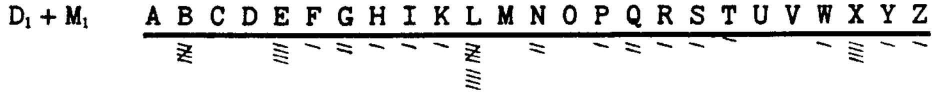
Applying the χ-test, the observed value of χ=34, the expected value is .05 (23×16)=18. An excellent match is obtained, and the hypothesis that D and M are in the same row in the checkerboard seems promising. Can any confirmation be found in the cryptogram itself?

i. It has already been pointed out that this system reduces to monoalphabetic substitution with variants. This being the case it should be possible to find manifestations of equivalency between some of the variant forms of $\frac{\Theta_1}{\Theta_2}$ vertical pairs in the cryptogram. If the student will think over the matter he will quickly see that this manifestation of equivalency is but a reflection of the principle elucidated in paragraph 46, expressed in a little different way. In other words, establishing equivalence between two Θ_1 components means that the two base letters involved belong in the same row of the checkerboard; establishing equivalence betweeen two Θ_2 components means that the two base letters involved belong in the same column of the checkerboard. Note the following instances of apparent equivalency between D_1 and M_1:

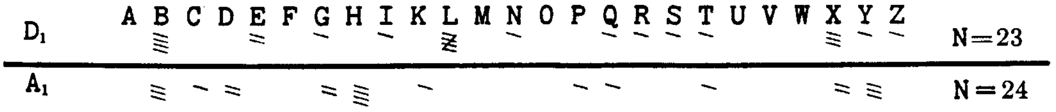
Period		
16	$D_2 B_1$	$D_1 B_2$
20	$Y_2 B_1$	$D_1 B_2$
49	$T_2 B_1$	$D_1 B_2$
2	$Y_2 B_1$	$M_1 B_2$
3	$Z_2 G_1$	$D_1 G_2$
56	$N_2 G_1$	$M_1 G_2$
13	$Q_2 L_1$	$D_1 L_2$
37	$Z_2 L_1$	$D_1 L_2$
58	$N_2 L_1$	$D_1 L_2$
66	$Z_2 L_1$	$D_1 L_2$
71	$Z_2 L_1$	$D_1 L_2$
6	$A_2 L_1$	$M_1 L_2$
13	$Y_2 L_1$	$M_1 L_2$
58	$I_2 L_1$	$M_1 L_2$

Period			
18	$G_2 Q_1$	$D_1 Q_2$	
32	$L_2 Q_1$	$M_1 Q_2$	
16	$F_2 E_1$	$D_1 E_2$	
17	$F_2 E_1$	$D_1 E_2$	
59	$G_2 E_1$	$M_1 E_2$	
12	$A_1 B_2$	$A_2 D_1$	$B_1 D_2$
50	$A_1 B_2$	$A_2 D_1$	$H_1 D_2$
8	$D_1 B_2$	$D_2 M_1$	$R_1 M_2$
19	$D_1 N_2$	$D_2 D_1$	$U_1 D_2$
46	$D_1 I_2$	$D_2 D_1$	$N_1 D_2$
44	$D_1 Z_2$	$D_2 M_1$	$B_1 M_2$
43	$U_1 A_2$	$U_2 D_1$	$C_1 D_2$
67	$U_1 E_2$	$U_2 M_1$	$P_1 M_2$

It may be assumed $D_1=M_1$ and the two distributions in figure 93 may be amalgamated.



The only other row in figure 90 which gives indications of being similar to this distribution is the A row. Applying the χ-test individually to the D_1 and M_1 distributions, and then to the combined D_1+M_1 distribution:



Expected for plain text: .05 (23×21)=24
Expected for random text: .038 (23×21)=18
Observed =26

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
M_1		/			//	/	/	/		/	////		/		/	/						/	/			N=16
A_1		///	/	//			//	////		/					/	/			/				//	///		N=21

Expected for plain text: .05 (16×21)=17
Expected for random text: .038 (16×21)=13
Observed =14

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
$D_1 M_1$		/////			////	/	//	/	/	/	///// ////		//		/	//	/	/	/			/	////	/	/	N=39
A_1		///	/	//			//	////		/					/	/			/				//	///		N=21

Expected for plain text: .05 (39×21)=41
Expected for random text: .038 (39×21)=31
Observed =39

From the foregoing calculations it appears that A_1 may be similar to the D_1 and the M_1 distributions, for the observed values, while not as great as expected for plain text, are higher than those expected for random text. Perhaps more conclusive evidence may be found if a search were made through the text to see if any equivalences between A_1, D_1, and M_1, appear.

Note the following cases:

Period			
8	$Y_2 B_1$	$A_1 B_2$	$A_2 M_1$
12	$N_2 B_1$	$A_1 B_2$	$A_2 D_1$
50	$N_2 B_1$	$A_1 B_2$	$A_2 D_1$
16	$D_2 B_1$	$D_1 B_2$	$D_2 A_1$
20	$Y_2 B_1$	$D_1 B_2$	$D_2 A_1$
49	$T_2 B_1$	$D_1 B_2$	$D_2 F_1$
2	$Y_2 B_1$	$M_1 B_2$	$M_2 R_1$
21	$C_2 G_1$	$A_1 G_2$	$A_2 X_1$
30	$K_2 G_1$	$A_1 G_2$	$S_1 P_2$
3	$Z_2 G_1$	$D_1 G_2$	$B_1 K_2$
56	$N_2 G_1$	$M_1 G_2$	$M_2 R_1$
24	$X_2 K_1$	$A_1 K_2$	$R_1 Z_2$
34	$T_2 K_1$	$M_1 K_2$	$M_2 Q_1$

Period			
33	$L_2 P_1$	$A_1 P_2$	$A_2 W_1$
48	$U_2 P_1$	$M_1 P_2$	$O_1 M_2$
15	$G_2 Q_1$	$A_1 Q_2$	$A_2 B_1$
18	$G_2 Q_1$	$D_1 Q_2$	$D_1 N_2$
32	$L_2 Q_1$	$M_1 Q_2$	$M_2 H_1$
14	$N_2 X_1$	$A_1 X_2$	$A_2 C_1$
61	$F_2 X_1$	$D_1 X_2$	$D_2 H_1$
57	$Q_2 Y_1$	$A_1 Y_2$	$I_1 D_2$
72	$F_2 Y_1$	$A_1 Y_2$	$A_2 D_1$
63	$U_2 Y_1$	$D_1 Y_2$	$D_2 A_1$

It certainly seems as though $A_1 = D_1 = M_1$, and that these letters are in the same row in the checkerboard. This tentatively will be assumed to be correct.

j. Among the most frequent combinations is the pair $Y_2 B_1$, appearing in the following sequences:

Period 2	$Y_2 C_1$	$Y_1 C_2$	$Y_2 B_1$	$M_1 B_2$	$M_2 R_1$
8	$L_1 Q_2$	$Y_1 W_2$	$Y_2 B_1$	$A_1 B_2$	$A_2 M_1$
10	$B_1 A_2$	$Y_1 O_2$	$Y_2 B_1$	$E_1 B_2$	$E_2 R_1$
20	$N_1 L_2$	$Y_1 W_2$	$Y_2 B_1$	$D_1 B_2$	$D_2 A_1$
41	$H_1 L_2$	$Y_1 M_2$	$Y_2 B_1$	$S_1 B_2$	$S_2 O_1$

Note how M_1, A_1, E_1, D_1, and S_1 all appear to be interchangeable. Are these the 5 letters which belong in the same row? The probable equivalence among A_1, D_1, and M_1 has been established by noting cases of equivalency in the text. A further search will be made to see if E_1 and S_1 also show equivalencies with A_1, D_1, and M_1.

Note the following:

Period 21	$C_2 G_1$	$A_1 G_2$	
30	$K_2 G_1$	$A_1 G_2$	
3	$Z_2 G_1$	$D_1 G_2$	
69	$O_2 G_1$	$E_1 G_2$	
56	$N_2 G_1$	$M_1 G_2$	
23	$D_1 X_2$	$D_2 H_1$	$A_1 H_2$
32	$M_1 Q_2$	$M_2 H_1$	$A_1 H_2$
61	$D_1 X_2$	$D_2 H_1$	$E_1 H_2$

Period 12	$N_2 B_1$	$A_1 B_2$	$A_2 D_1$
10	$Y_2 B_1$	$E_1 B_2$	$E_2 R_1$
30	$L_2 P_1$	$E_1 P_2$	$E_2 R_1$
33	$L_2 P_1$	$A_1 P_2$	$A_2 W_1$
43	$C_2 Y_1$	$E_1 Y_2$	$I_1 L_2$
57	$Q_2 Y_1$	$A_1 Y_2$	$I_1 D_2$

Here are indications that E_1 belongs to the same series, but not enough cases where S_1 is interchangeable with A, D, E, or M can be found to be convincing. But perhaps it is best not to go too fast in these early stages. Let it be assumed for the present that A, D, E, and M are in the same row of the substitution checkerboard. In period 16 there is the pair of vertical components $D_1 E_2$. Since $D_1 = E_1$ this pair may be written $E_1 E_2$, whereupon the plain-text letter E is immediately indicated. All cases of this sort are sought in the text and the plain-text letters are inserted in their proper places, there being 7 such instances in all, but these yield the important letters, A, D, and E.

k. In a similar manner, by an intensive search for cases in which components appear to be equivalent because they occur in repetitions which are identical save for one or two components, it is established that C, O, M, and W are in the same column in the checkerboard. Note the bracketing of these letters occurring as Θ_2 components in the 4th column of the first list of sequences in subparagraph *j.* Likewise, B, H, and N are established as being in the same row. Again the text is examined for cases in which plain-text letters C, O, M, W, B, H, and N may be inserted. By carrying out this process to the full extent possible, the skeletons of words will soon begin to appear.

l. Enough has been demonstrated to show this line of attack. Of course, if there is a large volume of text at hand, the simplest procedure would be to construct frequency distributions of the types shown in figures 90 and 91, and use the statistical method to match the individual distributions. For this method to be reliable it would be necessary to have several hundred letters of text, but this in actual practice would not be too much to expect.

m. There is, however, another line of attack, based upon the probable-word method. It has been pointed out that, in the case of letters in odd positions in the periods, 40 percent of the time the plain-text letter involved is indicated by either its Θ_1 or Θ_2 component. This property affords a fair basis for assuming a probable word. For example, the cryptogram here studied shows the following two periods:

Period				35							36			
Plain text					S			L						
Components	V_1	V_2	H_1	H_2	S_1	S_2	I_1	L_1	L_2	W_1	W_2	X_1	X_2	L_1
	I	X_1	X_2	S_1	S_2	L_1	L_2	L_2	X_1	X_2	R_1	R_2	S_1	S_2
Cipher text	V	H	S	I	X	S	L	L	W	X	L	X	R	S

Two letters are quite definite, S_s and L_s. Suppose the possible plain-text letters be indicated.

Period				35							36			
Possible plain-text letters	V	.	H	.	S	.	I	L	.	W	.	X	.	L
	I		X				L			X		R		S
Components	V_1	V_2	H_1	H_2	S_1	S_2	I_1	L_1	L_2	W_1	W_2	X_1	X_2	L_1
	I_2	X_1	X_2	S_1	S_2	L_1	L_2	L_2	X_1	X_2	R_1	R_2	S_1	S_2
Cipher text	V	H	S	I	X	S	L	L	W	X	L	X	R	S

The word HOSTILE is suggested by the letters H . S . I L . . This word will be assumed to be correct and it will be written out with its components under the cipher components. Thus:

Period				35				36	
Plain text			H	O	S	T	I	L	E
Cipher-text components	V_1	V_2	H_1	H_2	S_1	S_2	I_1	L_1	L_2
	I_2	X_1	X_2	S_1	S_2	L_1	L_2	L_2	X_1
Plain-text components			H_1	O_1	S_1	T_1	I_1	L_1	E_1
			H_2	O_2	S_2	T_2	I_2	L_2	E_2

This word, if correct, yields the following equivalencies: $H_2=X_2=O_1$; $S_1=O_2$; $T_1=S_2$; $L_1=T_2$; $I_2=L_2=E_1$; $X_1=E_2$. Again the text is examined for cases in which the plain-text letters may now be directly inserted; but only one case is found, in period 44, where $I_1 L_2=I_1 I_2=I_p$. This is unfortunate, so that additional words will have to be assumed. The 14th period shows a C_p and the components after it suggest that the word CROSSROADS may be present. Thus:

Period				14							15			
Plain text					C	R	O	S	S	R	O	A	D	S
Components	N_1	N_2	A_1	A_2	C_1	C_2	O_1	R_1	R_2	G_1	G_2	A_1	A_2	S_1
	O_2	X_1	X_2	C_1	C_2	R_1	R_2	S_2	W_1	W_2	Q_1	Q_2	B_1	B_2
Cipher text	N	A	C	O	X	C	R	R	G	A	S	W	Q	B

Take the first letter R_p, represented by $C_2 R_1$.

Since $R_p=C_2 R_1$,
Therefore, $R_1 R_2=C_2 R_1$
Hence $R_1=C_2$ and $R_2=R_1$
Therefore, $R_1=R_2=C_2$

Again, in the case of the first O_p,

$O_p=O_1 R_2$
But $O_p=O_1 O_2=O_1 R_2$
$O_2=R_2$
Therefore, $R_1=R_2=O_2=C_2$

The various equivalencies yielded are as follows:

$C_2=R_1=O_2=S_1=R_2=G_1=W_2=Q_1$
$S_2=W_1=B_2=T_1$
$H_2=X_2=O_1=G_2$
$O_2=Q_1$
$Q_2=A_2=D_1$

$B_1=D_2$
$L_1=T_2$
$L_2=I_2=E_1$
$X_1=E_2$

n. Let all the equivalencies found thus far from subparagraphs *e*, *f*, and *h* be collected in two tables, as shown in figure 94, one for Θ_1 Θ_2 combinations, the other for Θ_2 Θ_1 combinations.

	A_1	B_1	C_1	D_1	E_1	F_1	G_1	H_1	I_1	K_1	L_1	M_1	N_1	O_1	P_1	Q_1	R_1	S_1	T_1	U_1	V_1	W_1	X_1	Y_1	Z_1
$\Theta_1\Theta_2$ COMBINATIONS	D_1	H_1		A_1	A_1		O_2	B_1			T_2	A_1	B_1	H_2		G_1	G_1	O_2	S_2			T_1	E_2		
	E_1	N_1		E_1	D_1		C_2	N_1				D_1	H_1	X_2		O_2	C_2	C_2	W_1			S_2			
	M_1	D_2		M_1	M_1		M_2	D_2				E_1	D_2	G_2		C_2	O_2	M_2	B_2			B_2			
	I_2			I_2	I_2		W_2					I_2				M_2	M_2	W_2							
	L_2			L_2	L_2		R_1					L_2				W_2	W_2	R_1							
	A_2			A_2	A_2		R_2					A_2				R_1	R_2	R_2							
	Q_2			Q_2	Q_2		S_1					Q_2				R_2	S_1	G_1							
							Q_1									S_1	Q_1	Q_1							

	A_2	B_2	C_2	D_2	E_2	F_2	G_2	H_2	I_2	K_2	L_2	M_2	N_2	O_2	P_2	Q_2	R_2	S_2	T_2	U_2	V_2	W_2	X_2	Y_2	Z_2
$\Theta_2\Theta_1$ COMBINATIONS	A_1	T_1	M_2	B_1	X_1		O_1	X_2	L_2		I_2	C_2		C_2		L_2	C_2	T_1	L_1			C_2	H_2		
	D_1	S_2	O_2	H_1			H_2	O_1	A_1		A_1	O_2		M_2		I_2	M_2	W_1				M_2	O_1		
	E_1	W_1	W_2	N_1			X_2	G_2	D_1		D_1	W_2		W_2		A_1	O_2	B_2				O_2	G_2		
	M_1		R_1						E_1		E_1	R_1		S_1		D_1	W_2					S_1			
	I_2		S_1						M_1		M_1	S_1		R_1		E_1	S_1					R_1			
	L_2		R_2						A_2		A_2	R_2		R_2		M_1	R_1					R_2			
	Q_2		G_1						Q_2		Q_2	G_1		G_1		A_2	G_1					G_1			
			Q_1									Q_1		Q_1			Q_1					Q_1			

FIGURE 94

A study of the equivalencies indicates that—

(1) A, D, E, M belong in the same row.
(2) B, H, N belong in the same row.
(3) G, R, S, Q belong in the same row.
(4) R, C, O, M, W belong in the same column.
(5) I, L, A, Q belong in the same column.
(6) X, H, G belong in the same column.
(7) The coordinates of R and A are identical and hence this letter occupies a cell along a diagonal sloping from left to right in the matrix.

o. Since a row or a column can contain only 5 letters, it is obvious that A, D, E, M; B, H, N; and G, R, Q, S, fall in 3 different rows; C, O, M, W, R and I, L, A, Q fall in different columns. A start may be made by an arbitrary placement of R in the position 1–1, and since $R_1=O_2=C_2=$

$M_2=W_2$, this means that R, O, C, M, and W form one column in the substitution checkerboard, as shown in figure 95–A. The data also indicate that R, G, Q, and S must be in row 1, A, D, and E must be in row 4, H and X must be in column 3. This means that θ_1 for A, D, and E must be 4, and that θ_2 for H and X must be 3. And since $M_1=I_2=L_2$, θ_2 for I and L must be 4. Substituting in the text the coordinates for the known values, additional plain-text words soon become evident. For example, taking the periods with the word HOSTILE, it becomes possible to insert the letters R_p and O_p as the second and fourth letters after E_p, suggesting that the word after HOSTILE is TROOP. This gives $W_1 X_2=T_p$, which permits of placing T in position 5–3. Since T in HOSTILE$=S_2 L_1$, therefore $S_2=5$ and $L_1=3$. Since S is in row 1, and $S_2=5$, S must go in position 1–5. Since $L_2=4$ and $L_1=3$, L must go in position 3–4. Since O_p (the 1st O in TROOP)$=X_1R_2$ and it is known that $O_p=3–1$, therefore X must be in position 3–3. The checkerboard is now as shown in figure 95–B. From figure 94, $X_1=E_2$. Now $X_1=3$, and since the E must be in row 4,

2d component

1st component	1	2	3	4	5	
1	R					GS
2	C					
3	O					X
4	M					ADE
5	W					
			H X	I L		

FIGURE 95–A.

Period	35		36
Plain text	H O S T I		L E
Components	$V_1 V_2 H_1 H_2 S_1 S_2 I_1$		$L_1 L_2 W_1 W_2 X_1 X_2 L_1$
	$I_2 X_1 X_2 S_1 S_2 L_1 L_2$		$L_2 X_1 X_2 R_1 R_2 S_1 S_2$
Cipher text	V H S I X S L		L W X L X R S

θ_2

θ_1	1	2	3	4	5	
1	R				S	G
2	C					
3	O		X	L		
4	M					ADE
5	W		T			
			H	I		

FIGURE 95–B.

θ_2

θ_1	1	2	3	4	5	
1	R				S	G
2	C					
3	O		X	L		
4	M		E			AD
5	W		T			
			H	I		

FIGURE 95–C.

it is evident that E must occupy cell 4–3, as seen in figure 95–C. There are now only 2 possible rows for H, either 1 or 2. It is deemed unnecessary to give further details of the process. Suffice it to say that in a few minutes the entire checkerboard is found to be as shown in figure 95–D. It will decipher the entire cryptogram as it stands, but speculating upon the presence of W U T V Z in the last row, and assuming a key-word mixed sequence has brought this about, a rearrangement of the columns of the checkerboard is made to give T U V W Z, as shown in figure 95–E. The arrangement of the rows now becomes quite evident and the original checkerboard is found to be as shown in figure 95–F. It seems to be based upon the key phrase XYLOPHONIC BEDLAM.

Θ_2

	1	2	3	4	5
1	R	K	G	Q	S
2	C	N	H	I	B
Θ_1 3	O	Y	X	L	P
4	M	D	E	A	F
5	W	U	T	V	Z

FIGURE 95–D.

Θ_2

	3	2	4	1	5
1	G	K	Q	R	S
2	H	N	I	C	B
Θ_1 3	X	Y	L	O	P
4	E	D	A	M	F
5	T	U	V	W	Z

FIGURE 95–E.

Θ_2

	1	2	3	4	5
1	X	Y	L	O	P
2	H	N	I	C	B
Θ_1 3	E	D	A	M	F
4	G	K	Q	R	S
5	T	U	V	W	Z

FIGURE 95–F.

p. The completely deciphered cryptogram is as follows:

```
      1                2                3                4
S I T U A T I    O N O N F R O    N T O F T W E    N T Y F O U R
4 2 5 5 3 5 2    1 2 1 2 3 4 1    2 5 1 3 5 5 3    2 5 1 3 1 5 4
5 3 1 2 3 1 3    4 2 4 2 5 4 4    2 1 4 5 1 4 1    2 1 2 5 4 2 4
K Z F B E I L    Y Y M O C B R    B L Z D O T G    B L P K Y W C
            7               14               21               28

      5                6                7                8
T H B R I G A    D E A S F O L    L O W S C O L    O N F I R S T
5 2 2 4 2 4 3    3 3 3 4 3 1 1    1 1 5 4 2 1 1    1 2 3 2 4 4 5
1 1 5 4 3 1 3    2 1 3 5 5 4 3    3 4 4 5 4 4 3    4 2 5 3 4 5 1
U C C E P Q L    A M E Y L Z Q    X W H L R W Q    Y D R W B M T
           35               42               49               56

      9               10               11               12
B A T T A L I    O N F O R T Y    S E V E N T H    I N F A N T R
2 3 5 5 3 1 2    1 2 3 1 4 5 1    4 3 5 3 2 5 2    2 2 3 3 2 5 4
5 3 1 1 3 3 3    4 2 5 4 4 1 2    5 1 3 1 2 1 1    3 2 5 3 2 1 4
I Z E B E L A    Y E S O B R Y    Q V B B L Y X    N A B Q B D O
           63               70               77               84

     13               14               15               16
Y H A S R E A    C H E D C R O    S S R O A D S    E V E N F I V
1 2 3 4 4 3 3    2 2 3 3 2 4 1    4 4 4 1 3 3 4    3 5 3 2 3 2 5
2 1 3 5 4 1 3    4 1 1 2 4 4 4    5 5 4 4 3 2 5    1 3 1 2 5 3 3
Y M Q D L W L    N A C O X C R    R G A S W Q B    F D D T E B A
           91               98              105              112
      17                  18                  19                  20
E S E V E N D       D A S H R O A       D J U N C T I       O N F I V E T
3 4 3 5 3 2 3       3 3 4 2 4 1 3       3 2 5 2 2 5 2       1 2 3 2 5 3 5
1 5 1 3 1 2 2       2 3 5 1 4 4 3       2 3 2 2 4 1 3       4 2 5 3 3 1 1
M F D E T E N       A K G D F O Q       D U B N D C L       Y D V W B A X
          119                 126                 133                 140

      21                  22                  23                  24
H R E E T H R       E E G S T O P       E N E M Y H O       L D S W O O D
2 4 3 3 5 2 4       3 3 4 4 5 1 1       3 2 3 3 1 2 1       1 3 4 5 1 1 3
1 4 1 1 1 1 4       1 1 1 5 1 4 5       1 2 1 4 2 1 4       3 2 5 4 4 4 2
C A U G G X O       A R T X X T S       D A Y X H K O       L S X A B R K
          147                 154                 161                 168

      25                  26                  27                  28
S S O U T H W       E S T O F C H       A R L E S T O       W N I N C O N
4 4 1 5 5 2 5       3 4 5 1 3 2 2       3 4 1 3 4 5 1       5 2 2 2 2 1 2
5 5 4 2 1 1 4       1 5 1 4 5 4 1       3 4 3 1 5 1 4       4 2 3 2 4 4 2
R P U Z W H O       M T D H T S G       M L S L Q P O       U N H C I C K
          175                 182                 189                 196

      29                  30                  31                  32
S I D E R A B       L E F O R C E       S T O P W I L       L M A K E E V
4 2 3 3 4 3 2       1 3 3 1 4 2 3       4 5 1 1 5 2 1       1 3 3 4 3 3 5
5 3 2 1 4 3 5       3 1 5 4 4 4 1       5 1 4 5 4 3 3       3 4 3 2 1 1 3
K A Q B D O F       L E K A P R G       S X U P O W A       L M A V Q H L
          203                 210                 217                 224

      33                  34                  35                  36
E R Y E F F O       R T T O D R I       V E H O S T I       L E T R O O P
3 4 1 3 3 3 1       4 5 5 1 3 4 2       5 3 2 1 4 5 2       1 3 5 4 1 1 1
1 4 2 1 5 5 4       4 1 1 4 2 4 3       3 1 1 4 5 1 3       3 1 1 4 4 4 5
M L A X K P W       S T M C X K Q       V H S I X S L       L W X L X R S
          231                 238                 245                 252

      37                  38                  39                  40
S O U T A N D       O C C U P Y D       E F E N S I V       E P O S I T I
4 1 5 5 3 2 3       1 2 2 5 1 1 3       3 3 3 2 4 2 5       3 1 1 4 2 5 2
5 4 2 1 3 2 2       4 4 4 2 5 2 2       1 5 1 2 5 3 3       1 5 4 5 3 1 3
G Z D F K L N       Y B X M R B N       A D K T T B A       E O B H W V L
          259                 266                 273                 280

      41                  42                  43                  44
O N S T O P M       Y T R O O P S       H A V I N G D       I F F I C U L
1 2 4 5 1 1 3       1 5 4 1 1 1 4       2 3 5 2 2 4 3       2 3 3 2 2 5 1
4 2 5 1 4 5 4       2 1 4 4 4 5 5       1 3 3 3 2 1 2       3 5 5 3 4 2 3
Y S X M B O W       P G X K O R Z       I U C E A D Y       I D B L Z M I
          287                 294                 301                 308

      45                  46                  47                  48
T Y M A I N T       A I N I N G C       O N N E C T I       O N W I T H F
5 1 3 3 2 2 2       3 2 2 2 2 4 2       1 2 2 3 2 5 2       1 2 5 2 5 2 3
1 2 4 3 3 2 3       3 3 2 3 2 1 4       4 2 2 1 4 1 3       4 2 4 3 1 1 5
T A N H C A I       D N C I D D O       Y I B C N O L       Y U U M C E P
          315                 322                 329                 336
```

49 O R T Y F I F 1 4 5 1 3 2 3 4 4 1 2 5 3 5 O T D M G B F 343	50 T H I N F A N 5 2 2 2 3 3 2 1 1 3 2 5 3 2 U N A H L B D 350	51 T R Y O N N O 5 4 1 1 2 2 1 1 4 2 4 2 2 4 W X N X K K C 357	52 R T H S T O P 4 5 2 4 5 1 1 4 1 1 5 1 4 5 S C T O X T S 364
53 E N E M Y N O 3 2 3 3 1 2 1 1 2 1 4 2 2 4 D A Y X H K C 371	54 N C O M M I S 2 2 1 3 3 2 4 2 4 4 4 4 3 5 N L D K R R F 378	55 S I O N E D O 4 2 1 2 3 3 1 5 3 4 2 1 2 4 K Y A P M H C 385	56 F F I C E R C 3 3 2 2 3 4 2 5 5 3 4 1 4 4 A N M B V G R 392
57 A P T U R E D 3 1 5 5 4 3 3 3 5 1 2 4 1 2 E Z Q A T C Y 399	58 N E A R C H A 2 3 3 4 2 2 3 2 1 3 4 4 1 3 I M N D L R L 406	59 R L E S T O W 4 1 3 4 5 1 5 4 3 1 5 1 4 4 G M T W E T R 413	60 N S T A T E S 2 4 5 3 5 3 4 2 5 1 3 1 1 5 C V V K T E P 420
61 T H A T E N E 5 2 3 5 3 2 3 1 1 3 1 1 2 1 U F D E L X H 427	62 M Y S E V E N 3 1 4 3 5 3 2 4 2 5 1 3 1 2 E Q V C B L Y 434	63 T H D I V I S 5 2 3 2 5 2 4 1 1 2 3 3 3 5 U D U G Y A F 441	64 I O N I S M O 2 1 2 2 4 3 1 3 4 2 3 5 4 4 H N Q L K F R 448
65 V I N G I N T 5 2 2 4 2 2 5 3 3 2 1 3 2 1 U C N V D L H 455	66 O A T T A C K 1 3 5 5 3 2 4 4 3 1 1 3 4 2 L Z D R E L K 462	67 P O S I T I O 1 1 4 2 5 2 1 5 4 5 3 1 3 4 X K U P S E M 469	68 N S T O N I G 2 4 5 1 2 2 4 2 5 1 4 2 3 1 C T N K T K E 476
69 H T P R E P A 2 5 1 4 3 1 3 1 1 5 4 1 5 3 B O E E P G V 483	70 R A T O R Y T 4 3 5 1 4 1 5 4 3 1 4 4 2 1 Q T G W E R H 490	71 O A T T A C K 1 3 5 5 3 2 4 4 3 1 1 3 4 2 L Z D R E L K 497	72 A T D A Y L I 3 5 3 3 1 1 2 3 1 2 3 2 3 3 F A X I Y D A 504
73 G H T T O M O 4 2 5 5 1 3 1 1 1 1 1 4 4 4 K Z L X X O R 511	74 R R O W M O R 4 4 1 5 3 1 4 4 4 4 4 4 4 4 R P E R R R R 518	75 N I N G 2 2 2 4 2 3 3 1 N C I E 522	

q. The steps taken in recovering the original substitution checkerboard demonstrate that cyclic permutations of a correct checkerboard will serve to decipher such a cryptogram just as well as the original checkerboard. In other words, a cryptogram prepared according to this method is decipherable by factorial 5 (5×4×3×2×1=120) checkerboards, all of which are cyclically equivalent. Even though the identities of the components will be different if the same message is enciphered by two different cyclically-equivalent checkerboards, when these components are recombined, they will yield identical cipher texts, and therefore so far as external appearances are concerned different checkerboards yield identical cryptograms. The reason that there are only factorial 5 cyclically-equivalent checkerboards and not factorial 10, is that whatever permutation is applied to the row coordinates must be the same as that applied to the column coordinates in order that the aforesaid relationship hold true. If two checkerboards have identical row coordinates but different column coordinates certain portions of the cryptographic text will decipher correctly, others incorrectly. For this reason, in working with cryptograms of this type the cryptanalyst may successfully use a checkerboard which is incorrect in part and correct it as he progresses with the solution. It may also be added that the actual permutation of digits applied to the side and top of the checkerboard is of no consequence, so long as the permutations are identical. In other words, the permutation 5–2–1–3–4 will work just as well as 3–2–4–1–5, or 1–2–3–4–5, etc., so long as the same permutation is used for both row and column coordinates. It is the order of the rows and columns in the checkerboard which is the determining element in this system. Any arrangement (of the letters within the checkerboard) which retains the original order as regards the letters within rows and columns will work just as well as the original checkerboard.

r. A final remark may be worth adding. After all, the security of cryptograms enciphered by the bifid fractionating method rests upon the secrecy inherent in a 25-cell matrix containing a single mixed alphabet. In ordinary substitution, a single mixed alphabet hardly provides any security at all. Why does the bifid system, which also uses only a single mixed alphabet, yield so much higher a degree of security? Is it because of the transpositional features involved? Thinking about this point gives a negative answer, for after all, finding the length of the periods and replacing the cryptographic text by components based upon the cipher letters is a relatively easy matter. The transpositional features are really insignificant. No, the answer to the question lies in a different direction and may be summed up about as follows. In solving a simple mixed-alphabet substitution cipher one can attack a few cipher letters (the ones of greatest frequency) and find their equivalents, yielding fragments of good plain text here and there in the cipher text. Once a few values have been established in this manner, say 6 values, the remaining 20 values can be found almost from the context alone. And in establishing these 6 values, the letters involved are not so interrelated that all 6 have to be ascertained simultaneously. *The cryptanalyst may establish the values one at a time.* But in the case of the bifid system the equivalents of the plain-text letters are so interrelated that the cryptanalyst is forced to assume or establish the positions of several letters in the checkerboard *simultaneously*, not one by one. In other words, to use an analogy which may be only partially justified, the solution of a simple monoalphabetic substitution cipher is somewhat like forcing one's way into an inner chamber which has a number of doors each having a single lock; the solution of a bifid fractionated cipher is somewhat like getting into a vault—there is only one door which is provided with a complex 5-combination lock and all the tumblers of the lock must be positioned correctly *simultaneously* before the releasing lever can drop into the slot and the door opened. Fundamentally, this principle is responsible for the very much greater security of the bifid system as compared with that afforded by the simple monoalphabetic system. It is a principle well worth remembering and speculating upon.

53. Special solutions for bifid systems.—*a.* The security of the bifid system is very considerably reduced if the situation in which it is employed happens to be such that two or more messages with identical beginnings, endings, or internal portions can often be expected to occur. For in this case it is possible to establish equivalencies between components and quickly reconstruct the substitution checkerboard. An example will be given to illustrate the steps in a specific case.

b. Here are two cryptograms transmitted by two coordinate units to a superior headquarters at about the same time. They show certain identities, which have been underlined.

No. 1. QVBBL YXNAB QBDOY HONDW VUYTE MHQZD QTLKE EWAPK QSLIP QDWC
No. 2. VBNHY XDABG BDOIH OBNWV LYTFW HQXDQ VLKEW WAXDQ SABCA NXGX

c. Apparently these two cryptograms contain almost identical texts. In order to bring the identities into the form of superimposed components, it is necessary to transcribe the texts into periods of 7 and to superimpose the two messages as shown in figure 96.

d. The shifting of the second cryptogram 2 intervals to the right brings about the superimposition of the majority of θ_1 and θ_2 components and it may be assumed that for the most part the texts are identical. Allowing for slight differences at the beginnings and ends of the two messages, suppose a table of equivalencies is drawn up, beginning with the eighth superimposed pairs. Thus, $\frac{N_1}{Q_2}=\frac{N_2}{D_1}$; hence $N_1=N_2$ and $Q_2=D_1$. $\frac{N_2}{B_1}=\frac{H_1}{D_2}$; hence $N_2=H_1$ and $B_1=D_2$. Going through the text in this manner and terminating with the 42d superimposed pairs, the results are tabulated as shown in figure 97.

e. From these equivalencies it is possible to reconstruct, if not the complete substitution matrix, then at least a portion of the matrix. For example, the data show that N, H, B, and I belong in the same row; E and F belong in the same row; N, D, U, Y, and K belong in the same column, and so on. Experimentation to make all the data fit one checkerboard would sooner or later result in reconstructing the checkerboard shown in figure 95–F, and the two messages read as follows:

1. SEVENTH INFANTRY IN POSITION TO ATTACK AT FOUR AM PLAN FOUR.
2. TENTH INFANTRY IN POSITION TO ATTACK AT FOUR AM PLAN THREEX.

f. The foregoing gives a clue to what would happen in the case of an extensive traffic in which long phrases or entire sentences may be expected to occur repeatedly. By a proper indexing of all the material, identical sequences would be uncovered and these, attacked along the lines indicated, would soon result in reconstructing the checkerboard, whereupon all the messages may be read with ease.

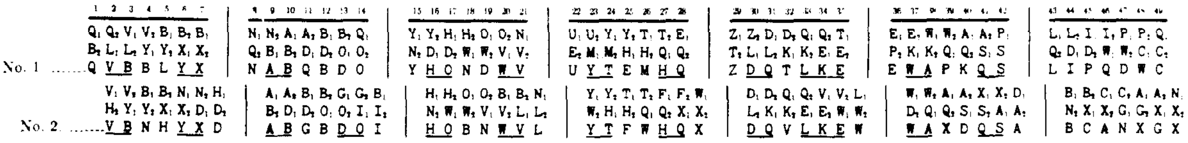
54. Solution of trifid systems.—*a.* In the trifid fractionating system the cipher alphabet is tripartite in nature, that is, the plain-text letters are represented by permutations of 3 components taken in groups of 3's, thus forming a set of 27 equivalents, such as that shown below:

A=111
B=112
C=113
D=121
E=122
F=123
G=131
H=132
I=133

J=211
K=212
L=213
M=221
N=222
O=223
P=231
Q=232
R=233

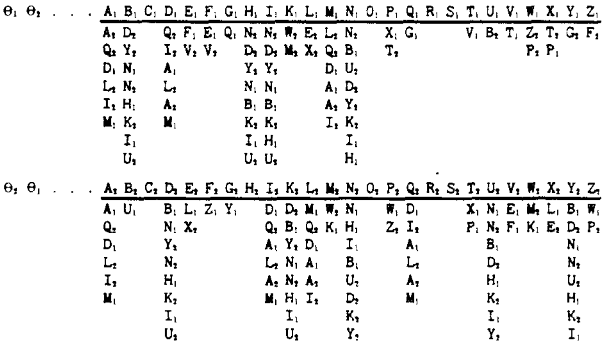
S=311
T=312
U=313
V=321
W=322
X=323
Y=331
Z=332
?=333

b. The equivalents may, of course, be arranged in a mixed order, and it is possible to use one tripartite alphabet for decomposition and a wholly different one for recomposition. One disadvantage of such an alphabet is that it is a 27-element alphabet and therefore some subterfuge must be adopted as regards the 27th element, such as that illustrated in the footnote to paragraph 57 of Special Text No. 166, *Advanced Military Cryptography,* wherein ZA stands for Z and ZB for the 27th character.



	1 2 3 4 5 6 7	8 9 10 11 12 13 14	15 16 17 18 19 20 21	22 23 24 25 26 27 28	29 30 31 32 33 34 35	36 37 38 39 40 41 42	43 44 45 46 47 48 49
	$Q_1\ Q_2\ V_1\ V_2\ B_1\ B_2\ B_1$	$N_1\ N_2\ A_1\ A_2\ B_1\ B_2\ Q_1$	$Y_1\ Y_2\ H_1\ H_2\ O_1\ O_2\ N_1$	$U_1\ U_2\ Y_1\ Y_2\ T_1\ T_2\ E_1$	$Z_1\ Z_2\ D_1\ D_2\ Q_1\ Q_2\ T_1$	$E_1\ E_2\ W_1\ W_2\ A_1\ A_2\ P_1$	$L_1\ L_2\ I_1\ I_2\ P_1\ P_2\ Q_1$
	$B_2\ L_1\ L_2\ Y_1\ Y_2\ X_1\ X_2$	$Q_2\ B_1\ B_2\ D_1\ D_2\ O_1\ O_2$	$N_2\ D_1\ D_2\ W_1\ W_2\ V_1\ V_2$	$E_2\ M_1\ M_2\ H_1\ H_2\ Q_1\ Q_2$	$T_2\ L_1\ L_2\ K_1\ K_2\ E_1\ E_2$	$P_2\ K_1\ K_2\ Q_1\ Q_2\ S_1\ S$	$Q_2\ D_1\ D_2\ W_1\ W_2\ C_1\ C_2$
No. 1	$Q\ \underline{V\ B}\ B\ L\ \underline{Y\ X}$	$N\ \underline{A\ B}\ Q\ B\ D\ O$	$Y\ \underline{H\ O}\ N\ D\ \underline{W\ V}$	$U\ \underline{Y\ T}\ E\ M\ \underline{H\ Q}$	$Z\ \underline{D\ Q}\ T\ \underline{L\ K\ E}$	$E\ \underline{W\ A}\ P\ K\ \underline{Q\ S}$	$L\ I\ P\ Q\ D\ W\ C$
	$V_1\ V_2\ B_1\ B_2\ N_1\ N_2\ H_1$	$A_1\ A_2\ B_1\ B_2\ G_1\ G_2\ B_1$	$H_1\ H_2\ O_1\ O_2\ B_1\ B_2\ N_1$	$Y_1\ Y_2\ T_1\ T_2\ F_1\ F_2\ W_1$	$D_1\ D_2\ Q_1\ Q_2\ V_1\ V_2\ L_1$	$W_1\ W_2\ A_1\ A_2\ X_1\ X_2\ D_1$	$B_1\ B_2\ C_1\ C_2\ A_1\ A_2\ N_1$
	$H_2\ Y_1\ Y_2\ X_1\ X_2\ D_1\ D_2$	$B_2\ D_1\ D_2\ O_1\ O_2\ I_1\ I_2$	$N_2\ W_1\ W_2\ V_1\ V_2\ L_1\ L_2$	$W_2\ H_1\ H_2\ Q_1\ Q_2\ X_1\ X_2$	$L_2\ K_1\ K_2\ E_1\ E_2\ W_1\ W_2$	$D_2\ Q_1\ Q_2\ S_1\ S_2\ A_1\ A_2$	$N_2\ X_1\ X_2\ G_1\ G_2\ X_1\ X_2$
No. 2	$\underline{V\ B}\ N\ H\ \underline{Y\ X}\ D$	$\underline{A\ B}\ G\ B\ \underline{D\ O}\ I$	$\underline{H\ O}\ B\ N\ \underline{W\ V}\ L$	$\underline{Y\ T}\ F\ W\ \underline{H\ Q}\ X$	$\underline{D\ Q}\ V\ \underline{L\ K\ E}\ W$	$\underline{W\ A}\ X\ D\ \underline{Q\ S}\ A$	$B\ C\ A\ N\ X\ G\ X$

Figure 96.



$\Theta_1\ \Theta_2$ . . .	A_1	B_1	C_1	D_1	E_1	F_1	G_1	H_1	I_1	K_1	L_1	M_1	N_1	O_1	P_1	Q_1	R_1	S_1	T_1	U_1	V_1	W_1	X_1	Y_1	Z_1
	A_2	D_2		Q_2	F_1	E_1	Q_1	N_2	N_2	W_2	E_2	L_2	N_2		X_1	G_1			V_1	B_2	T_1	Z_2	T_2	G_2	F_2
	Q_2	Y_2		I_2	V_2	V_2		D_2	D_2	M_2	X_2	Q_2	B_1		T_2							P_2	P_1		
	D_1	N_1		A_1				Y_2	Y_2			D_1	U_2												
	L_2	N_2		L_2				N_1	N_1			A_1	D_2												
	I_2	H_1		A_2				B_1	B_1			A_2	Y_2												
	M_1	K_2		M_1				K_2	K_2			I_2	K_2												
		I_1						I_1	H_1				I_1												
		U_2						U_2	U_2				H_1												

$\Theta_2\ \Theta_1$ . . .	A_2	B_2	C_2	D_2	E_2	F_2	G_2	H_2	I_2	K_2	L_2	M_2	N_2	O_2	P_2	Q_2	R_2	S_2	T_2	U_2	V_2	W_2	X_2	Y_2	Z_2
	A_1	U_1		B_1	L_1	Z_1	Y_1		D_1	D_2	M_1	W_2	N_1		W_1	D_1			X_1	N_1	E_1	M_2	L_1	B_1	W_1
	Q_2			N_1	X_2				Q_2	B_1	Q_2	K_1	H_1		Z_2	I_2			P_1	N_2	F_1	K_1	E_2	D_2	P_2
	D_1			Y_2					A_1	Y_2	D_1		I_1			A_1				B_1				N_1	
	L_2			N_2					L_2	N_1	A_1		B_1			L_2				D_2				N_2	
	I_2			H_1					A_2	N_2	A_2		U_2			A_2				H_1				U_2	
	M_1			K_2					M_1	H_1	I_2		D_2			M_1				K_2				H_1	
				I_1						I_1			K_2							I_1				K_2	
				U_2						U_2			Y_2							Y_2				I_1	

Figure 97

c. The various types of fractionation possible in bifid systems are also adaptable in trifid systems. For example, using the alphabet shown above for recomposition as well as decomposition the encipherment of a message in periods of 7 is as follows:

Plain text	R	E	L	I	E	F	O	F	Y	O	U	R	R	E	G	I	M	E	N	T	T	O	M	O	R	R	O	W
Components	2	1	2.	1	1	1.	2	1	3	2.	3	2	2.	1	1	1	2.	1	2	3.	3	2	2	2.	2	2	2.	3
	3	2.	1	3	2.	2	2	2	3.	2	1	3.	3	2	3	3.	2	2	2.	1	1	2	2.	2	3	3.	2	2
	3.	2	3	3.	2	3	3	3.	1	3	3.	3	3	2	1.	3	1	2.	2	2	2	3.	1	3	3.	3	3	2
Cipher text	K	A	Q	H	O	R	R	H	W	F	L	X	I	Z	B	F	?	N	A	T	N	N	N	W	R	O	I	Z

CRYPTOGRAM

KAQHO RRHWF LXIZA BFZBN ATNNN WROIZ

d. The solution of a single cryptogram of this nature would be a quite difficult matter, especially if there were nothing upon which to make assumptions for probable words. But a whole series of cryptograms could be solved, following in general the procedure outlined in the case of the bifid system, although the solution is, admittedly, much more complicated. The first step is to ascertain the length of the period, and when this has been done, transcribe the cipher text into components, which in their vertical combinations then represent monoalphabetic equivalents, with, of course, many variants for each letter of the plain text. Then a study is made to establish component equivalents, just as in the bifid system. If the text is replete with repetitions, or if a long word or a short phrase may be assumed to be present, a start may be made and once this sort of entering wedge has been forced into the structure, its further disintegration and ultimate complete demolition is only a matter of time and patience.

55. Concluding remarks on fractionating systems.—*a*. It goes without saying that the basic principles of fractionation in the bifid and trifid systems are susceptible to a great deal of variation and complication. For example, instead of having periods of fixed length through the message it is possible to vary the length of the periods according to some simple or complex key suitable for this purpose. Or the bifid and trifid systems may be combined into a single scheme, enciphering a text by the bifid method and then reenciphering the cipher text by the trifid method and so on. Systems of this sort may become so complex as to defy analysis, especially if the keys are constantly and frequently varied so that no great amount of traffic accumulates in any single key. Fortunately for the cryptanalyst, however, such complex systems as these, if introduced into actual usage, are attended by so many difficulties in practice that the enemy cryptographic service would certainly break down and it would not be long before requests for repetition, the transmission of the same cryptogram in different keys, and so on, would afford clues to solution. Could such systems be employed successfully in field service there is no doubt that from the standpoint of security, the cryptograms would be theoretically secure. But the danger of error and the slowness with which they could be operated by the usual cryptographic clerks are such that systems of this complexity can hardly be employed in the field, and therefore the cryptanalyst may not expect to encounter them.

b. However, the simple bifid system, the ADFGVX system, and the like, are indeed practicable for field use, have been used with success in the past, and may be expected to be in use in the future. It is therefore advisable that the student become thoroughly familiar with the basic principles of their solution and practice the application of these principles as frequently as possible. In this connection, the attention of the student is directed to the fact that there is theoretically no reason why the bipartite components of the ADFGVX system cannot be recombined by means of the same or a different checkerboard, thus reducing the cryptographic text to a form wherein it consists of 25 different letters, and at the same time cutting the length of the messages in half. The matter is purely one of practicability: it adds one more step to the process. But it must not be overlooked that this additional step would add a good deal of strength to the system, for it would shorten, mask, distort, or entirely eliminate similar beginnings and similar endings—the two most fruitful sources of attack on this system.

56. Concluding remarks on transposition systems.—*a*. Simple transposition systems hardly afford any security at all; complex ones may in the case of individual or single messages afford a high degree of security. But just as soon as many cryptograms in the same key are transmitted the chances of finding two or more cryptograms of identical length become quite good and the general solution may be applied.

b. Contrary to the situation in the case of substitution, in that of transposition wherein the letters of the plain-text itself are transposed (not code) the shorter the cryptogram the greater the possibility of solution. For, in the case of a message of say only 25 or 30 letters, one might shift the letters about and actually reconstruct the plain text as one does in the case of the game called "anagrams." Of course, several different "solutions" may thus be obtained, but having such "solutions" it may be possible to reconstruct the system upon which the transposition was based and thus "prove" one of the solutions.

c. The text has confined itself almost entirely to cases of uniliteral transposition, in order to demonstrate basic principles. But there is inherently no reason why transposition may not be applied to digraphs, trigraphs, or tetragraphs. If longer sequences are used as the units of transposition the security decreases very sharply, as in the case of the ordinary route ciphers of the Civil War period.

d. Transposition designs, diagrams, or patterns are susceptible of yielding cryptograms of good security, if they are at all irregular or provide for nulls and blank spaces. Such devices are particularly difficult to solve if frequently changed.

e. Transpositions effected upon fixed-length sequences of plain text yield a low degree of security but when a transposition is applied to the cipher text resulting from a good substitution system or to the code text of cryptograms first encoded by means of an extensive code book the increase in the cryptographic security of such cryptograms is quite notable. In fact, transposition methods and designs are frequently used to "superencipher" substitution text or code and play a very important role in this field. Their great disadvantage is that inherent in all transposition methods: The addition or deletion of a single letter or two often makes the entire cryptogram unreadable even with the correct key.

f. The clues afforded by messages with similar beginnings, endings, or internal portions, and by repetitions of incorrectly enciphered messages without paraphrasing the original text are often sufficient to make a solution possible or to facilitate a solution. For this reason the cryptanalyst should note all cases wherein clues of this sort may be applicable and be prepared to take full advantage of them.

SECTION XI

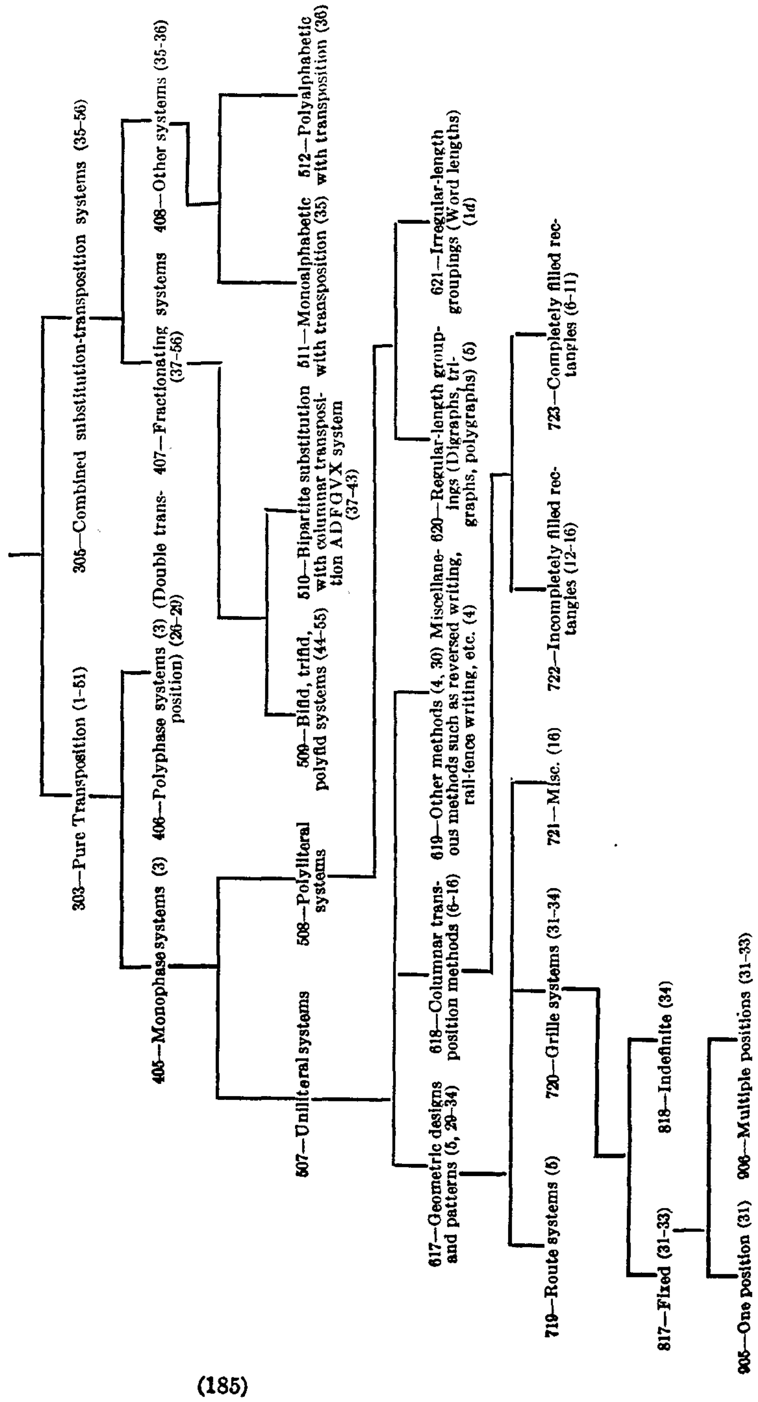
ANALYTICAL KEY

Paragraph
Analytical key 57

57. Analytical key.—Continuing the scheme initiated in the first text of this series, the analytical key applicable to the subject matter and systems embraced in this text is given below.

ANALYTICAL KEY FOR MILITARY CRYPTANALYSIS, IV

(Numbers in parenthesis refer to paragraph numbers in this text)


Transposition Systems
303—Pure Transposition (1-51)
305—Combined substitution-transposition systems (35-56)
405—Monophase systems (3)
406—Polyphase systems (3) (Double transposition) (26-29)
407—Fractionating systems (37-56)
408—Other systems (35-36)
507—Unilteral systems
508—Polyliteral systems
509—Bifid, trifid, polyfid systems (44-55)
510—Bipartite substitution with columnar transposition ADFGVX system (37-43)
511—Monoalphabetic with transposition (35)
512—Polyalphabetic with transposition (36)
617—Geometric designs and patterns (5, 29-34)
618—Columnar transposition methods (6-16)
619—Other methods (4, 30) Miscellaneous methods such as reversed writing, rail-fence writing, etc. (4)
620—Regular-length groupings (Digraphs, trigraphs, polygraphs) (5)
621—Irregular-length groupings (Word lengths) (1d)
719—Route systems (5)
720—Grille systems (31-34)
721—Misc. (16)
722—Incompletely filled rectangles (12-16)
723—Completely filled rectangles (6-11)
817—Fixed (31-33)
818—Indefinite (34)
905—One position (31)
906—Multiple positions (31-33)

INDEX

Page

- **ADFGVX** system, 97–143
  - General solution, 124–143
    - Alternation of components, 125–126
    - Basic principles, 125–127, 130–131
    - Final components, 125–126
    - Illustration of solution, 127–143
    - Initial components, 125–126
    - Minus alternation, 131
    - Plus alternation, 131
  - Special solution, 98–124
    - Exact-factor method of solution, 123–124
    - Solution by means of identical beginnings, 105–123
    - Solution by means of identical endings, 98–105
- Alternation of components, 110, 131
- Anagram sequence, 33
- Anagramming, 5, 51–53
- Analytical key, 185
- Ascertaining period in bifid system, 165–166

- Bifid fractionating systems, 144–183
  - Ascertaining period, 165–166
  - Basic steps of, 144–146–160–161
  - Bipartite equivalents of, 144–147
  - Column coordinate, 146
  - Even-length periods, 161
  - General principles underlying solution, 146–150, 160–165
  - Illustration of solution, 150–160, 166–181
  - Matching of distributions, 170–173
  - Odd-length periods, 161
  - Periods of fixed length, 160
  - Preparation of index, 151, 166
  - Probability of occurrence of even-positioned letters, 170–171
  - Probability of occurrence of odd-positioned letters, 170–171
  - Probable-word method of attack, 174–176
  - Reconstruction of original substitution checkerboard, 152–154, 175–178
  - Row coordinate, 146
  - Security of, 181, 183
  - Solution of, 146–183
  - Special solution of, 156–160, 181–182
  - Vertical pairs of components, 144–145
- Bipartite equivalents, 144, 162
- Blanks (in matrix), 31, 184
- "Breaks", 35–36
- Break table, 36, 100
- "Cage", 2, 4
- Cipher→Plain sequence, 32, 55–75

Page

- Columnar transposition ciphers, 3–79
  - Completely-filled rectangles, 4–17
    - Column and row transposition, 17
    - Consonants and vowels, deviation of, 6–10
    - Invariable digraph, 14
    - Keyword reconstruction diagram, 15
    - Limited affinity, 14
    - Matrix, 2
    - Matrix reconstruction, 80–84
    - Obligatory sequences, 14
    - Pilot letters, 14
    - Probable-word method of solution, 13–14, 37–39
    - Reconstruction of literal key, 15–17, 25
  - General solution, 18, 51–53
  - Incompletely-filled rectangles, 18–36
    - Alternative method of solution, 25–31
    - Formula for calculating length and number of long and short columns, 18
    - General principles underlying solution, 18–24
    - General solution, 18, 51–53
    - Keyword reconstruction diagram, 15, 25
    - Long columns of, 18
    - Short columns of, 18
    - Special solution of, 37–38, 40–55
    - Width, 18
  - Special solutions, 37–38, 40–55
    - Cryptograms of identical length in same key, 51–53
    - Interchanged pair of columns, 43–44
    - Messages with similar beginnings, 44–47
    - Messages with similar endings, 47–49
    - Omitted column, 42
    - Single message containing a long repetition, 49–50
    - Stereotyped phraseology, 37–39
- Combined substitution-transposition systems, 94–96
  - Using digraphic substitution, 96
  - Using fractionating systems, 96
  - Using known alphabets, 94–95
  - Using monoalphabetic substitution, 94–95
  - Using polyalphabetic or polygraphic substitution, 96
- Completely-filled rectangles, 4
- C→P sequence, 32, 55–75
- "Crown" diagram, 20–21
- Cyclic permutation of transposition key, 35, 180

- Double transposition ciphers, 51–79
  - Depth of rectangle a multiple of width, 78–79
  - Enciphering rectangle a perfect square, 76
  - Failure to execute double transposition properly, 75–76
  - Reconstructing keys, 55–75
  - Special cases of solution, 75–79
  - Width of rectangle a multiple of depth, 76–78

- Encipher sequence, 33
- Exact-factor method of solving ADFGVX cipher, 123–124

Page
Fractionating systems 97–184
  ADFGVX 97–143
  Bifid 144–182
  Trifid 182–183
"Frame" 2, 4

General solution 32, 37–38, 51–53, 124–125
Geometric designs 80–84
Grilles, indefinite or continuous 91–93
Grilles, revolving 85–91
  Alpha method 85–91
  Beta method 85
  Principle of exclusion 89
  Principle of sequence 89
  Principle of symmetry 87

"Hat" diagram 20–21

Inscription 2
Interchanged pair of columns 43–44
Interval sequence 58
Invariable digraph 14
Invariant relationship 69
Inverse sequence 32–33

Keyword reconstruction diagram 15
κρ sequence 33

Limited affinity 14
Literal key, reconstruction of 15–17, 25
Logarithms of probabilities, use of 6, 12–13, 143

Matching distributions 170–173
Matrix 2, 4
Matrix reconstruction 80–84
Monophase transposition 2

Nulls (in matrix) 1, 31, 184

Obligatory sequences 14
Omitted column 42

Partial C→P sequence 75
P→C interval sequence 58
P→C sequence 32, 55–75
Pilot letters 14
Polyphase transposition 2
Processes, rescriptive 2
Plain→Cipher sequence 32, 55–75

Rail-fence writing 3
Rescription, process of 2
Reversed writing 3, 95
Route transposition 3–4

Page

Single transposition 2
Solution by superimposition 51–53
Special solution 37–38, 40–55, 75–79
Superimposition, solution by 51–53

Term number 34
Transcription 2
Transposition:
  Columnar 4
  Double 55–79
  Monophase 2
  Polyphase 2
  Sequence 33
  Simple types of 3–17
  Single 2
  Uniliteral route 3–4
  Vertical writing 3
Trifid fractionating system, solution of 182–183

Uniliteral transposition 4

Vertical writing 3

κρ sequence 33

BOOKS IN THE CRYPTOGRAPHIC SERIES

C-1	MANUAL FOR THE SOLUTION OF MILITARY CIPHERS, Parker'Hitt
c-2	CRYPT'ALYSIS OF THE SIMPLE SUBSTITUTION CIPHER, Wayne G. Barker
c-3	ELEMENTS OF CRYTANALYSIS, William F. Friedman
C-4	STATISTICAL METHODS IN CRYPTANALYSIS, Solomon Kullback
C-5	CRYPTOGRAPHY AND CRYPTANALYSIS ARTICLES, Vol 1, Friedman
C-6	CRYPTOGRAPHY AND CRYPTANALYSIS ARTICLES, Vol 2, Friedman
c-7	ELEMENTARY MILITARY CRYPTOGRAPHY William F. Friedman
C-8	ADVANCED MILITARY CRYPTOGRAPHY, William F. Friedman
c-11	SOLVING GERMAN CODES IN WORLD WAR I, William F. Friedman
c-14	MANUAL OF CRYPTGGRAPHY, Luigi Sacco
c-17	CRYPTANALYSIS OF THE HAGELIN CRYPTOGRAPH, Wayne G. Barker
c-20	HISTORY OF CODES AND CIPHERS IN THE U.S. PRIOR TO WORLD WAR I, cd. Barker
c-21	HISTORY OF CODES AND CIPHERS IN THE U.S. DURING WORLD WAR I, ed. Barker
c-22	HISTORY OF CODES AND CIPHERS IN THE U.S. DURING THE PERIOD BETWEEN THE WORLD WARS, PART I, 1919-1929, ed. Barker
C-26	CRYPTANALYSIS OF AN ENCIPHERED CODE PROBLEM, Wayne G. Barker
c-30	MILITARY CRYPTANALYSIS, PART I, William F. Friedman
c-33	COURSE IN CRYPTANALYSIS, Volume 1, British War Office
C-34	COURSE IN CRYPTANALYSIS, Volume 2, British War Office
c-35	THE ORIGIN AND DEVELOPMENT OF THE NATIONAL SECURITY AGENCY, George A. Brownell
C-36	TREATISE ON CRYPTOGRAPHY, Andre Lange and S.A. Soudart
c-39	CRYPTANALYSIS OF SHIFT-REGISTER GENERATED STREAM CIPHER SYSTEMS, Barker
C-46	MILITARY CRYPT'ALYSIS, PART II, William F. Friedman
c-41	ELEMENTARY COURSE IN PROBABILITY FOR THE CRYPTANALYST, Andrew M. Gleason
c-42	MILITARY CRYPTANALYTICS, PART I, VOL. 1, Friedman & Callimahos
c-43	MILITARY CRYPTANALYTICS, PART I, VOL. 2, Friedman & Callimabos
C-44	MILITARY CRYPTANALYTICS, PART II, VOL. 1, Callimahos & Friedman
C-45	MILITARY CRYPT'ALYTICS, PART II, VOL. 2, Callimahos & Friedman
C-46	PATTERN WORDS – THREE LETTERS TO EIGHT LETTERS IN LENGTH, Carlisle
C-48	PATTERN WORDS – NINE LETTERS IN LENGTH, Sheila Carlisle
c-49	THE INDEX OF COINCIDENCE AND ITS APPLICATIONS IN CRYPTANALYSIS, Friedman
c-so	CRYPTOGRAPHIC SIGNIFICANCE OF THE KNAPSACK PROBLEM, O'Conner and Seberry
c-52	THE AMERICAN BLACK CHAMBER, Herbert 0.. Yardley
c-53	TRAFFIC ANALYSIS AND THE ZENDIAN PROBLEM, L.D. Callimahos
C-54	HISTORY OF CODES AND CIPHERS IN THE U.S. DURING THE PERIOD BETWEEN THE WORLD WARS, PART II, 1930-1939, ed. Barker
C-55	INTRODUCTION TO THE ANALYSIS'OF THE DATA ENCRYPTION STANDARD (DES), Barker
c-54	ELEMENTARY CRYPTOGRAPHY AND CRYPTANALYSIS, Donald D. Millikin
C-57	SECRETS CIPHERS OF THE 1876 PRESIDENTIAL ELECTION, D. Beaird Glover
C-58	SOLVING CIPHER PROBLEMS, Frank W. Lewis
c-59	CRYPTANALYSIS OF THE SINGLE COLUMNAR TRANSPOSITION CIPHER, W. Barker
C-60	MILITARY CRYPTANALYSIS, PART III, William F. Friedman
C-61	MILITARY CRYPTANALYSIS, PART IV, William F. Friedman
C-62	PATERN WORDS – TEN-LETTERS AND ELEVEN-LETI'ERS IN LENGTH, Wallace
C-63	PATTERN WORDS – TWELVE-LE'ITERS AND GREATER IN LENGTH, Wallace
C-64	U.S. NAVAL CRYPTOGRAPHIC ACTIVITIES IN THE PHILIPPINES PRIOR TO WWII, ed. Carlisle
c-65	U.S. NAVAL COMMUNICATIONS INTELLIGENCE ACTMTIES, Safford & Wenger

ISBN 0-89432-398-7

90000

9 780894 121982

ISBN: O-8941 2-I